REDUCING CRIME AND VIOLENCE:
EXPERIMENTAL EVIDENCE ON ADULT NONCOGNITIVE INVESTMENTS IN LIBERIA

Christopher Blattman
Julian C. Jamison
Margaret Sheridan

Reducing Crime and Violence: Experimental Evidence on Adult Noncognitive Investments
in Liberia
Christopher Blattman, Julian C. Jamison, and Margaret Sheridan
NBER Working Paper No. 21204
May 2015
JEL No. D03,J22,K42,O12

ABSTRACT

We show that self control, time preferences, and values are malleable in adults, and that investments
in these skills and preferences reduce crime and violence. We recruited criminally-engaged Liberian
men and randomized half to eight weeks of group cognitive behavioral therapy, fostering self regulation,
patience, and noncriminal values. We also randomized $200 grants. Cash alone and therapy alone
dramatically reduced crime and violence, but effects dissipated within a year. When cash followed
therapy, however, crime and violence decreased by as much as 50% for at least a year. We hypothesize
that cash reinforced therapy's lessons by prolonging practice and self-investment.

Christopher Blattman
School of International and Public Affairs
Columbia University
420 West 118th Street
New York, NY 10027
and NBER
chrisblattman@columbia.edu

Julian C. Jamison
Office of Research
Consumer Financial Protection Bureau
1700 G Street NW
Washington, DC 20552
julison@gmail.com

Margaret Sheridan
University of North Carolina
Department of Psychology and Neuroscience
235 E. Cameron Avenue
Chapel Hill, NC
sheridan.margaret@unc.edu

A online appendix is available at http://www.nber.org/data-appendix/w21204

1 Introduction

In many countries, poor young men exhibit high rates of violence, crime, and other "anti-social" behaviors. In addition to their direct costs, crime and instability hinder economic growth by reducing investment or allocating resources to security. In fragile states, such men are also targets for mobilization into election intimidation, rioting, and rebellion.[1]

Two of the most common government responses are policing and job creation. Both take the person as they are and try to change their incentives or simply incarcerate them (Becker, 1968; Draca and Machin, 2015). This paper investigates an alternative: rehabilitation, or changing behavior by shaping people's underlying skills, identity, and values.

A large literature has shown that a broad set of noncognitive skills, especially self control, predict long-run economic performance and criminal activity.[2] These skills respond to investment, especially in childhood (Cunha et al., 2010). They are fostered by family, schools, and communities. There is little evidence, however, on the returns to late-stage noncognitive investments, and so it's unclear whether by adulthood self-investment or interventions can shape noncognitive skills and hence behavior (Heckman and Kautz, 2014; Hill et al., 2011). It's also unclear what specific skills are both important and malleable.

To investigate, we recruited 999 of the highest-risk men in Liberia's capital, generally aged 18 to 35. Most were engaged in part-time theft and drug dealing, and regularly had violent confrontations with each other, community members, and police.

We experimentally ran two interventions. One was an 8-week program of group cognitive behavior therapy (CBT) called the STYL program, for Sustainable Transformation of Youth in Liberia. We assigned offers by lottery. Following this, we held a second lottery for a $200 grant—about three months wages. The cash was partly a measurement tool, to see if therapy affected economic decisions. The cash was also a treatment, in the sense that it could stimulate legal self-employment, and we included it to try to compare therapy to a rise in the returns to legal work.[3] Experimentally, subjects either received offers of therapy, cash, therapy then cash, or neither. To deliver both treatments cost about $530 per person.

CBT is a therapeutic approach used to treat a range of harmful beliefs and behaviors, including depression, anger, and impulsivity. First, it tries to make people aware of and challenge harmful, automatic patterns of thinking or behavior. Second, it tries to disrupt these patterns of thinking and foster better ones by having people practice new skills and behaviors—learning by doing. A Liberian non-profit, the Network for Empowerment and

---

[1]For example, poor urban young men were recently recruited into election violence in Sierra Leone (Christensen and Utas, 2008) and as mercenaries in Cote d'Ivoire Blattman and Annan (2015).

[2]e.g. Nagin and Pogarsky, 2004; Heckman et al., 2006; Borghans et al., 2008

[3]Evidence from East Africa suggests that the poor and unemployed are credit-constrained and have high returns to cash (Haushofer and Shapiro, 2013; Blattman et al., 2014, 2015).

Progressive Initiatives (NEPI), designed and implemented STYL. NEPI facilitators were themselves ex-combatants or ex-criminals who graduated from a prior NEPI therapy.

Among "noncognitive skills," STYL focused foremost on self control. This includes more short term abilities to regulate one's emotions and be resistant to impulse, as well as more sustained abilities to be planful, persevering, and patient. Self control skills are central components of many programs, from preschool to rehabilitation therapy.[4] The curriculum focused on helping men foster skills of planning, goal-setting, reflective and deliberate decision-making, and controlling their emotions and impulses.

The therapy also tried to encourage nonviolent, noncriminal preferences by fostering a change in the men's self image. A premise of STYL was that the men self-identified as outcasts and did not hold themselves to the standards of mainstream society. The therapy tried to persuade the men that they could change who they were and how they were perceived. NEPI facilitators modeled this image change. They walked the men through basic steps, such as changing their appearance or engaging in normal social interactions. Therapy also required men to practice going to supermarkets, banks, and other "normal" places.

Literature in both psychology and economics supports the idea that self image and associated values influence behavior, and that both can change. This literature treats values as direct utility benefits or penalties from acting in accordance with or against a set of preferences (Bénabou and Tirole, 2004; Almlund et al., 2011). Akerlof and Kranton (2000) and Jolls et al. (1998) both argue that these preferences or values are tied to a person's self image, or perceived social category, and that to some extent people can change their social category and with it values that reward and penalize certain behaviors.

There are striking parallels between STYL and socialization into militaries, street culture, gangs and armed groups. Such groups use similar techniques (appearance change, practice, modeling) to shape young men's self-image and behavior (Vigil, 2003; Wood, 2008; Maruna and Roy, 2007). NEPI designed STYL to reverse this process.

We surveyed the men beforehand, a few weeks after the interventions, and a year later. Most had no fixed address, phone, or even name, and they moved frequently. Despite this mobility, we re-interviewed 93%. We rely on self-reported data since (like most poor and fragile states) there are no administrative data or arrest records. We did not necessarily trust self-reports, and so we validated behaviors such as drug use and stealing in a subsample.

We approached roughly 1500 high-risk men, and 999 agreed to enter the study. Of those assigned to therapy, nearly all attended at least a day, and two thirds completed it. The higher risk men were the most likely to finish.

---

[4]e.g. Gottfredson and Hirschi, 1990; Borghans et al., 2008. As an example of an intervention, the famous Perry Preschool Program emphasized the ability of young children to plan tasks, execute their plans, and review their work in social groups (Almlund et al., 2011).

Men who received therapy reduced their antisocial behavior dramatically. Within a few weeks, drug dealing halved and thefts fell by a third compared to controls. With therapy alone, these effects diminished after a year. When therapy was followed by cash, however, effects were lasting. A year later, those who received both therapy and cash were 44% less likely to be carrying a weapon, 43% less likely to sell drugs, and reported lower aggression. In the control group, men reported stealing almost once per week on average, and with therapy and cash this fell from 66 to 30 crimes per year per person.

Therapy probably worked through many channels, but we see evidence of improvement in all three channels we a priori theorized and pre-specified: self control skills, economic time preferences, and anticriminal values. With therapy alone, the noncognitive changes diminished after a year. When therapy was followed by cash, however, the effects of therapy on self control, time preferences, and anticriminal values were lasting and fairly large— at least 0.2 standard deviations. Treatment effects are similar whether we examine topics emphasized or not in the STYL curriculum.

How was cash used? Regardless of therapy, little of the grant was spent on drugs or "wasteful" things. Most funds were invested in business or saved. Cash led to a short-term increase in petty trading and income. After a year, however, these gains disappeared, partly because most men were robbed regularly, irrespective of treatment.

The fact that the cash grant was crucial to sustaining the effects of therapy is our most unexpected and important finding. Since we find no sustained effect of cash on earnings, cash clearly did not raise the opportunity cost of antisocial behavior after a year. Thus economic incentives do not explain the sustained effect of therapy plus cash on crime and aggression. Drawing on qualitative interviews and psychological theory, we suggest that the short term increase in income and legal employment helped to solidify therapy's impact on self control skills and values. Specifically, for a few months after therapy, cash allowed men to project a changed self, to stave off homelessness and stealing, and further practice the behavior change started by therapy. This hypothesis will be important to test in future research.

An obvious concern is our reliance on self-reported data. We argue that misreporting is unlikely to drive our results for two reasons. The first is the pattern of treatment effects: long term impacts from therapy plus cash, but not from cash or therapy only. Systematic measurement error would need to be correlated with the "both" treatment arm only. This seems feasible but unlikely, especially given the magnitudes of the impacts. To check further, we attempted to validate a subset of questions using intensive qualitative observation. The patterns suggest that, if anything, the control group underreported sensitive behaviors. Hence the treatment effects may actually underestimate therapy's impacts.

In addition to evaluating the pairing of an economic intervention with CBT, this study

addresses several important gaps in the literature. The most obvious is the absence of evidence outside the U.S., and the importance of such evidence in fragile states.

Even within the U.S., however, there is limited evidence on adult behavior change. Most program evaluations focus on education and employment interventions.[5] Studies of behavioral therapy tend to be small-sample and non-experimental (Wilson et al., 2005).[6]

Finally, few studies have measured noncognitive skill and value changes directly, and so our study strengthens arguments that self control and values are malleable and contribute to antisocial behavior. The malleability of self image is consistent with evidence from stigmatized Indian sex workers, where short courses of non-CBT psychological therapy increased self-worth, reduced shame, and increased savings and health-seeking behavior (Ghosal et al., 2015). The malleability of self control is also consistent with a large body of evidence in the U.S. showing that CBT programs in schools and correctional institutes reduce criminal recidivism, especially among adolescents and children.[7] The majority of this evidence, however, comes from small, observational, unpublished studies, which, because of a reliance on administrative data seldom measure mechanisms directly.[8] But three recent randomized control trials among at-risk Chicago adolescents show that CBT can help adolescents reduce automatic behaviors (such as violent retaliations to a slight) by learning to override "fast" decision-making with conscious "slow" reflection (Heller et al., 2013, 2015). The parallels between that program and STYL, in both the curriculum and impacts, are striking.

It remains to be seen if STYL is replicable elsewhere, but it is promising that STYL was adapted from foreign therapies, and also developed its own facilitators from prior graduates, enhancing scalability. Future work should test generalizability to new contexts and comparisons to other therapies (or a placebo), but also address the limitations of this study, including a reliance on self-reported data.

2 Intervention and experiment

Liberia's capital, Monrovia, is home to a third of the country's 4.3 million people. There are few formal jobs. Most men aged 18 to 35 have limited employment and earn money through a mix of agriculture, casual labor, or petty business. A few turn to crime, which is becoming

---

[5] Two U.S. programs, Job Corps and ChalleNGe, are residential programs for at-risk youth that tackle non-cognitive skills but focus on jobs and job training(Schochet et al., 2008; Millenky et al., 2012).

[6] An exception is Little et al. (1994), who randomly assigned CBT to 1,381 general offenders in Tennessee. They found that re-arrest fell from 56% to 41% after 5 years. Our study adds to this large-sample evidence.

[7] For evidence on children and adolescents, see Heckman and Kautz (2014); Hill et al. (2011). Meta-analyses of adolescent and adult interventions in correctional institutes find that CBT-informed programs appear to outperform alternate therapies (Andrews et al., 1990; Lipsey, 2009).

[8] Of 20 studies identified by Wilson et al. (2005) only four were experimental and three of these had sample sizes of 100 or less. The observational studies were also small and were of mixed quality.

more violent and commonplace.

From 1989-96 and 1999-2003 two civil wars wracked Liberia. They killed 10% of the population, displaced a majority, and recruited tens of thousands into combat. Since 2003, however, Liberia has been at peace with the help of a United Nations (UN) peacekeeping force. During our study period, 2009-12, the economy was growing 6% per year (Republic of Liberia, 2012). Nonetheless, in 2009, people aged 18 to 35 would have spent up to 15 years of their childhood or adolescence in conflict, many robbed of the institutions that normally fostered planfulness, emotional stability, and other noncognitive skills.

Marginalized young men are one of the government's main concerns, especially poorly integrated ex-combatants and other men involved in drugs and crime. Drug and criminal networks are disorganized, but the government worried they could consolidate. They also worried about political violence. High-risk men have joined riots and election violence in the past, and they were targets for mercenary recruitment into the 2010-11 war in Côte d'Ivoire.

2.1   Target population and recruitment

We set out to recruit 1000 high-risk men—men actively involved in crime, interpersonal violence, and drugs, or who were poor and at risk of engaging in these activities. With no administrative data on such men, we recruited them directly. We selected five neighborhoods in Monrovia known for high rates of crime. These were generally mixed-income residential areas with large markets, with populations of roughly 100,000.

Recruiters were NEPI affiliates who were not involved in the interventions. NEPI had extensive knowledge of these neighborhoods and connections to local leaders, as well as a strong reputation that target men could verify. Recruiters had worked closely with high risk men before, and were themselves past graduates of a NEPI program.

We charged the recruiters with finding men that were: homeless; drug-using; disreputable in appearance; or present in locations known for crime, armed recruitment, and violence. Location was especially important. Within each of the neighborhoods there were pockets of insecurity where high-risk men were known to live or congregate: abandoned buildings, garbage dumps, drug dealing spots, parking lots, and homes for rootless young men run by ex-military commanders. Community members could easily identify these spots and their denizens. Similarly, certain professions had strong reputations for crime.[9] Appearance was also a useful guide. For instance, recruiters looked for men with a dirty or unkempt appearance, long hair, apparent intoxication, or a "tough" style of dress.

---

[9]Professions included "car loaders" who have reputations for pickpocketing, or wheelbarrow and motorbike parking areas with reputations for drug selling and crime. They avoided recruiting men known to be "bosses"—men who run homes or drug dens that cater to petty criminals and low-level drug dealers.

To minimize correlated outcomes and spillovers, we avoided recruiting close associates. We instructed NEPI to approach just one out of every 7–10 high-risk men they visually identified. Recruiters then described the therapy, the allocation by lottery, and the baseline survey. They never mentioned cash grants.

Over several weeks, recruiters identified roughly 10,000 potentially high-risk men and approached 1,500. Of these, about one third refused interest in the therapy and survey.[10]

In the end, 999 men agreed to enter the sample. We estimate they represent 0.6% of all adult males in the neighborhoods, and about 12% of men aged 18–35 and in the bottom decile of income (Appendix A.2). Column 1 of Table 1 describes this sample at baseline. On average the men were 25, had nearly eight years of schooling, earned about \$68 in the past month working 46 hours per week (mainly in low skill labor and illicit work), and had \$34 informally saved. 38% were a former member of an armed group.

2.2   Interventions

Cash

A nonprofit organization, Global Communities (GC), distributed the cash. They ran a lottery, where winners received \$200 cash and losers received a consolation prize of \$10. There was minimal framing.[11] GC held cash lotteries a week after the end of therapy.

Therapy

CBT is a short-term approach that tries to reduce self-destructive beliefs or behaviors and promote positive ones. It does so in two ways. First, the therapist tries to help the patient become more aware of their automatic thoughts: inaccurate or negative thinking about themselves or others. Shifting automatic thoughts allows them to respond to everyday situations in a more effective way.

A central principle of CBT, however, is that sustained changes in behavior or symptoms also come from actively practicing new behaviors, often starting with simple tasks and, through repetition, positive reinforcement, and gradually increasing the difficulty or complexity of the tasks, changing both behavior and thinking. This practice happens in therapy

---

[10]We do not have systematic data on refusers, but recruiters reported two main types: men who were poor but were "low-risk" in that they did not appear to be involved in crime and violence; and high-risk men who said they were too busy to take part in therapy because they had legal or illegal business to attend to.

[11]See Appendix B.4 for implementation details. Prior to the lottery, subjects were given about 15 minutes of information on how to keep the money safe (e.g. depositing it with a bank) and examples of what they could use it for (e.g. starting a small business or home improvement). But GC explicitly emphasized to subjects that the grant was unconditional and they were free to do what they wished.

Table 1: Baseline summary statistics and test of balance for select covariates

	Sample	Test of randomization balance (N=999)						F-Test
		Assigned therapy		Assigned cash		Assigned both		
Baseline covariate	Mean	Coeff.	p value	Coeff.	p value	Coeff.	p value	p value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	25.40	-0.16	0.68	0.19	0.59	-0.18	0.68	0.18
Married or partnered	0.16	-0.03	0.65	-0.04	0.67	0.04	0.76	0.93
# children <15 in household	2.20	-0.59	0.07	-0.50	0.19	0.62	0.30	0.33
Years of schooling	7.72	-0.19	0.68	0.04	0.95	-0.01	0.99	0.55
Has any disabilities	0.08	0.04	0.29	0.00	1.00	-0.04	0.48	0.19
Ex-combatant	0.41	0.06	0.09	0.08	0.11	-0.09	0.12	0.15
Weekly cash earnings (USD)	17.02	-1.89	0.03	-4.85	0.03	5.48	0.00	0.02
Currently sleeping on street	0.25	-0.01	0.82	0.00	0.93	-0.02	0.74	0.33
Savings stock (USD)	33.83	-10.08	0.26	-12.74	0.31	15.71	0.31	0.53
Hours/week, illicit activities	13.55	1.21	0.68	-0.86	0.67	0.06	0.99	0.14
Hrs/week, agriculture	0.36	0.34	0.26	-0.10	0.35	0.13	0.84	0.01
Hrs/week, low-skill wage labor	19.39	0.54	0.88	1.24	0.73	-0.43	0.90	0.94
Hrs/week, in low-skill business	11.53	0.16	0.92	-1.53	0.60	5.76	0.13	0.50
Hrs/week, high-skill work	1.51	-0.05	0.91	0.94	0.03	0.11	0.85	0.01
Sells drugs	0.20	0.01	0.69	0.00	0.92	0.00	0.93	0.92
Uses marijuana daily	0.44	0.08	0.14	0.04	0.12	-0.09	0.21	0.34
Uses hard drugs daily	0.15	-0.04	0.21	0.02	0.52	0.01	0.90	0.37
Committed theft, past 2 weeks	0.53	0.05	0.51	0.01	0.64	-0.02	0.66	0.80
R-squared		0.17		0.11		0.33		
p value on F-statistic		0.53		0.66		0.25		

Notes: We report a selection of covariates here, and all 57 covariates are reported in Appendix A.1. Column (1) reports the sample mean. A small number of missing values are imputed at the median. Columns (2)-(7) report the coefficients and p values from ordinary least squares regressions of each baseline covariate on three indicators, one for assignment to each treatment arm, controlling for block fixed effects. Column (8) reports the p value from a joint test of statistical significance of all three treatment indicators.

but also as "homework" (Beck, 2011).[12]

Origins and aims of STYL STYL grew from of the experiences of NEPI's founders, but over time they integrated standard Western CBT practices, in part through interactions with international organizations and experts.

The program combined group therapy with one-on-one counseling. Twenty men met in groups three times a week, four hours at a time, led by two NEPI facilitators. On alternate days when groups did not meet, facilitators visited men at home or work to provide advising and encouragement. NEPI offered no compensation except lunch, since men who sacrificed

---

[12]CBT has been studied extensively and validated as a treatment for several of the behaviors targeted by STYL: anger, aggression, criminality, and substance abuse (Saini, 2009; Pearson et al., 2002; Wilson et al., 2005; Del Vecchio and O'Leary, 2004).

four hours of work could not afford to eat.

NEPI designed the curriculum and approach to encourage two kinds of change. First, they taught skills of self control: to manage anger and emotions, reduce impulsivity, become more conscientious and persevering, and become more forward-looking.[13] While often described as personality traits, such traits evolve over the life cycle and are affected by upbringing and investment, so we follow Heckman and Kautz (2014) in considering them skills of character. This concept of self control has close parallels to time preferences.[14] Our measures and model treat them as distinct, and whether they covary is an empirical question.

Second, NEPI tried to persuade men to adopt anti-criminal, anti-violent values by shifting their self image from outcast to normal society member. A premise of STYL was that the security and respect associated with a mainstream identity were familiar and attractive, to the men. So were the values associated with a mainstream identity—it was no mystery that crime and drugs were considered "bad". But those norms and values did not apply to outcasts like them, and a mainstream identity seemed out of reach.

NEPI facilitators tried to persuade the men that this identity was attainable, and that the men should at least try. Partly through exercising skills of self-control, and partly by practice and exposure to new situations, the STYL curriculum walked men through the process of change. The facilitators were an integral part of this intervention, because they modeled the change in skills and values. All were graduates of a prior STYL-like program run by NEPI, and three-quarters were former street youth or combatants.

There are parallels to interventions which show that aspirations—forward-looking goals or targets—influence behavior and respond to intervention (Bernard et al., 2014). There are also parallels to switching social identity.[15]

STYL curriculum and approach   The sessions employed a variety of techniques, from lectures and group discussions, to various forms of practice, including: role playing in class, homework that requires practicing tasks, exposure to real situations, and in-class processing of experiences of executing these tasks. Like many CBT programs, these tasks began simply

---

[13]Note that psychologists also use "self control" to refer to abilities such as executive function (EF) and delay of gratification (DoG), both of which are thought to lead to less impulsive decision-making and influence long-term success (Mischel et al., 1989). Some evidence suggests that EF and DoG are distinct from our character skills and are less malleable (Duckworth and Schulze, 2011). We measured EF and DoG but they were not the focus of the therapy and we did not hypothesize any change.

[14]In general, the literature is unclear whether character skills are related to time preferences. The limited evidence suggests correlations are positive but low (Becker et al., 2012).

[15]Akerlof and Kranton (2000) reviews a wide social science literature. Relatedly, criminologists sometimes refer to similar process of "knifing off" from old social rules and behaviors, and associate these changes with significant turning points in life, such as marriage, a move, or a life-threatening experience (Maruna and Roy, 2007). This literature almost always ties successful knifing off to having a new "script" for the future. The STYL program is effectively that script.

and got more difficult over time.[16]

In the first three weeks, facilitators encouraged men to try to maintain some new, simple behaviors. This included getting a haircut and removing facial hair, wearing shoes and pants instead of sandals and shorts, improving personal hygiene and the cleanliness of their living area, and reducing substance abuse. These simple exercises in goal-setting and self control also helped men start to operate within mainstream social norms.

In the middle weeks, facilitators encouraged men to engage with society in planned and unaccustomed ways, akin to exposure therapy.[17] For instance, homework included reintroducing themselves to their family, joining community sports, and visiting banks, supermarkets, shops, and so forth. Men also studied successful people in their community, and reached out to one as a mentor. Men then processed their attempts as a group. Often homework was independent, but facilitators might accompany the more troubled men.

Men also learned to manage emotion: practicing nonaggressive responses to angry confrontations in class, and recognizing signs of angry reactions and learning to distract or calm oneself (walking away, doing other activities, or breathing techniques).

In the last weeks, facilitators taught planning and goal setting. These lessons included training on breaking down large goals into smaller accomplishable sub-goals, and then creating plans to accomplish them via concrete steps. For example, men would list subgoals of a plan; these were written on a paper in front of the room, for all to see; the group critiqued them; and plans were rewritten. For homework men would attempt planning in their own lives: how to feed their family the next day; starting a garden; making a savings plan; reconciling with estranged family; or starting a business. These assignments began easy and got more difficult. This process of goal identification and planning is central to most forms of CBT, especially for disruptive behavior disorders (Langberg et al., 2013).

Cost The cost of delivering both interventions was $530 per head: $189 for CBT, $216 for the grant, and $125 for registration and administration.

2.3 Experimental design

We used a 2×2 factorial design. The experiment proceeded as follows: First, one week after recruitment and baseline surveys, NEPI held public draws to assign half the men to an offer of therapy in blocks of each day's recruits. Therapy began one week later. About 1–2 weeks

---

[16]Appendix B.3 describes the curriculum in more detail. The full program manual is available at http://chrisblattman.com/documents/policy/2015.STYL.Program.Manual.pdf.

[17]Therapy for patients with social phobia practice similar engagement (Ponniah and Hollon, 2008). Besides practice, subjects learn that social feedback is less negative than feared. By re-engaging with society, participants tested their negative beliefs about themselves.

after therapy, GC announced and held a private draw for \$200 grants among the full sample, in blocks of roughly 50 men. Finally, a third organization (Innovations for Poverty Action) ran endline surveys 2 and 5 weeks, and then 12 and 13 months, after grants.

The sample were very mistrustful of authority, and we randomized by individual draw rather than computerized assignment to maximize trust, transparency, and staff safety. Men in each block took turns drawing colored chips from a fabric bag.[18]

Balance This resulted in 25% assignment to cash only, 28% to therapy only, 25% to both, and 22% to neither (Table 2).[19] Treatment is balanced along covariates. Table 1 reports tests of balance for teach treatment, for selected covariates (see Appendix A.1 for all, and for endline respondents only). Of 57 covariates over three treatments, 14 (8.2%) have a difference with $p < .05$, and within treatment arms the covariates are not jointly significant.

Compliance Both interventions had high compliance, in part due to NEPI's persuasive efforts and street credibility. Of men assigned to the grant, 98% received it. Of men assigned to therapy, 5% attended none, another 5% dropped out within the first 3 weeks, and two thirds attended most sessions ($>80\%$) (Appendix A.4). Those who dropped out early had less schooling, less self control, and were less likely to exhibit antisocial behaviors like substance abuse or stealing (Appendix A.3). Thus the highest-risk men seem more likely to attend over poorer, noncriminal men.

Phased implementation For logistical reasons we recruited, treated, and studied the men in three phases, as seen in Table 2. A pilot phase recruited 100 men, to ensure that the therapy and cash grant caused no harm, to assess statistical power, and to allow us to refine experimental protocols. The pilot showed no indication of harm, and so we scaled to a further 900 with only minor changes to the interventions and protocols in two phases.[20]

3   Conceptual framework

Both interventions aim to reduce crime and violence. To structure our thinking and highlight key mechanisms, we start with a model of the effects of therapy and cash on economically-

---

[18]The order of selection was deliberately unsystematic but not randomized. The number of chips in the bag generally exceeded the number of draws, partly to avoid a correlation between order of the draw and treatment assignment probabilities, and partly to avoid having late-drawing men receive their status by default. See Appendix B.2 for full details.

[19]The excess therapy assignments is in part chance, and is in part driven by two blocks where excess treatment chips were accidentally used. All regressions include block fixed effects to account for this.

[20]Appendix B includes the power calculations behind our experimental design.

Table 2: Study sample and treatment assignment by block and phase

Phase	Start date (MM/YY)	Block (slum)	Sample	% recruits assigned to:			
				Therapy	Cash	Therapy & cash	Neither
1	12/10	Red Light	100	25.0%	25.0%	25.0%	25.0%
2	06/11	Red Light	219	26.9%	25.1%	24.2%	23.7%
	06/11	Central Monrovia	179	31.8%	19.0%	31.8%	17.3%
3	03/12	Clara Town	175	28.6%	27.4%	22.9%	21.1%
	02/12	Logan Town	86	26.7%	29.1%	19.8%	24.4%
	02/12	New Kru Town	240	26.3%	26.7%	23.8%	23.3%
All			999	27.7%	25.1%	24.9%	22.2%

motivated violence, such as crime or election thuggery. The simplest model treats crime as an occupational choice between legal and illegal work.[21] We then consider how such a model can capture more expressive and reactionary forms of anti-social behavior. We develop the formal model in Appendix C and outline the structure and results here.

The most unusual element here, of course, is CBT. NEPI's approach and the psychological theory underlying the therapy suggests it could influence criminal occupational choice in three main ways. First, improved emotional regulation, planning, and related noncognitive skills can be modeled as a type of human capital that enters the individual production function. Alternatively, therapy's effect on self control could be modeled as a change in time preferences, including time-inconsistency. Finally, we model a change in values as a change in intrinsic preferences over criminal occupations or other antisocial acts.[22]

Of course, the therapy is a multifaceted treatment that likely operates through a number of other mechanisms: changed peers, family reunification, skills of conflict resolution, reduced drug abuse, and so forth. We focus on (and had pre-specified) self control skills, time preferences, and antiviolent values because these mechanisms were most in line with NEPI's design principles and theories, as well as the psychological theory and evidence cited above.

Setup Suppose people can allocate their time between leisure l, legal work L^b such as petty business or labor, and illegal occupations L^c such as crime, mercenary work, or election thuggery. We refer to these simply as "business" and "crime".

We assume crime uses labor alone and pays a wage w, which may be uncertain. This resembles the returns we observe to illegal work of the type available to our population in

---

[21]It is rooted in models of occupational choice with capital infusions and adapted to illicit behavior, as in Blattman and Annan (2015), in the tradition of many economic crime models (Draca and Machin, 2015).

[22]Typically models treat such preferences as fixed, or ignore them. We outline how exogenous changes in noncognitive abilities or preferences affect the comparative statics in an otherwise standard model.

Liberia.[23] In the budget function, crime also carries a punishment f with probability ρ, and we assume this risk increases with the time devoted to crime. Punishment could mean prosecution, mob justice, or social sanctions.

Business uses capital, yielding output $F\left(\theta, L_t^b, K_t\right)$ where θ is individual ability and K_t is capital inputs. People start with wealth in the form of a riskless asset, a_0, and save or borrow at interest rate r. Self control skills are one element of θ, and output increases in θ.

Our model assumes that these self control skills are inputs into business but not crime. We did not assume this from the outset, recognizing that in principle STYL could teach men to be more effective criminals. The pilot phase, however, suggested the opposite was true.[24]

People choose consumption, labor supply in each sector, and the amount of wealth to invest in business (versus the safe asset) in order to maximize their utility subject to the constraint that consumption plus wealth are equal to total income from business, crime, and the interest on investment. We allow people to be present-biased in the sense that they have a general inter-temporal discount factor δ, but can also be time-inconsistent with an extra factor denoted $\beta < 1$ that multiplies all future periods relative to the present (the standard form of quasi-hyperbolic time preferences).

Finally, people value consumption and leisure, but we also allow for a consumption value from conforming to one's self-image and values (Akerlof and Kranton, 2000; Bénabou and Tirole, 2004).[25] In this case, a person's self image and values can penalize criminal acts. We use σ to indicate a preference against crime, and we put it in the utility function, $U\left(c, l, \sigma L^c\right)$, to distinguish these internal preferences from external punishments f.

We are interested in the effect of the interventions on criminal versus legal labor. Therapy can potentially influence this occupational choice through non-cognitive skills θ, time preferences (δ or β), anti-criminal values σ, or all of the above. Cash, meanwhile, can influence occupational choice by increasing the assets available for capital inputs into legal business.

Occupational choice in the absence of interventions Where financial markets work well and where people are time consistent ($\beta = 1$), businesses are at their optimal scale—they have borrowed until the marginal return to capital is equal to r. Of course, the poor are typically credit-constrained. In this case poor people are forced to invest in capital over time

---

[23]Petty crime requires little capital; drug dealers typically work for a "boss" who owns the supply; and those who leave town to work in illicit mining work as "mining boys" for capital-owning "miners" on short-term renewable contracts that pay a daily wage plus a payment tied to output. This is also why we assume below that self-control skills are less important for success in criminal activities.

[24]Also, this is the interesting and relevant case, since otherwise investments in self control skills will not affect occupational choice. Assuming that self control is an input into both sectors, but that the returns to it are higher in business, would deliver similar comparative statics.

[25]We ignore the possibility, proposed by Bénabou and Tirole (2004), that ability is imperfectly known and correlated with perceived self-image.

until they reach the same optimal scale. The young and those who have experienced bad shocks will be the furthest behind. As a result, crime is more likely to be chosen by men with low business ability θ, the poor and credit-constrained, those with low disutility of crime, and the time-inconsistent. People may also choose both crime and business. Credit-constrained people with partial capital for business may still spend some time in crime. Also, risk averse people may do both activities when returns are uncertain.

Impacts of cash If there are no credit constraints, cash windfalls will not affect occupational choice. But if people are poor and credit-constrained, windfalls will be partly invested in business. People involved in crime will shift to business, especially those with high business ability. Cash infusions will lead to a smaller increase in business for time-inconsistent individuals, however, since they will choose to consume more today.

Impacts of therapy Therapy could increase σ, θ, β or δ. These channels have distinguishing predictions. Interventions that increase the disutility from crime, σ, will reduce time devoted to it, but will have no effect on returns to business. Interventions that increase noncognitive ability θ will induce more time and investment in business, and also reduce crime. With the presence of risk in both sectors (and assuming risk aversion), interventions in θ will have relatively greater effects in terms of pushing individuals away from crime, because an increase in θ now also makes business relatively less risky. A rise in σ will also have a bigger effect than without uncertainty, because risk aversion will reinforce the rise in crime aversion and further reduce hours in crime.

What if an intervention increases time consistency, β? This will increase business investment and earnings among the credit-constrained. If people become more time-consistent, they will be more strongly influenced by the consequences of their actions in terms of punishments, and will therefore reduce criminal labor (and increase business labor) as well. Similar comparative statics come from an increase in patience.

Cash and therapy in combination Both interventions should lead to a larger decline than one alone simply because the effects are cumulative to a degree. Moreover, when people are credit-constrained and also receive cash, this simple model predicts that the effects of a change in σ or θ will be greater with cash than without it. Thus the interventions may be complementary and the total effect could be greater than the sum of the parts.

Relevance for aggression This model is most useful for thinking about pecuniary crime. On the other hand, some violence does not earn a wage and does not necessarily have an opportunity cost of time. Nonetheless, we can still use the model to think about such

aggression. For instance, we can think of some acts as having consumption value that is fleeting (the pleasure from expressing anger) or persistent (deterring future slights). In this case, $\sigma < 0$. Like crime, these acts carry a risk of future punishment.

If the criminal wage is zero, there is still a tradeoff between the consumption value today and the risk of punishment tomorrow, and the main comparative statics of therapy are similar to the case of crime: instilling values against violence (increasing σ) will reduce aggression; and increasing time consistency, β, also reduces aggression. Cash, however, will have little deterrent effect on aggression.[26]

4 Data

We tried to survey each subject five times: (i) at baseline prior to the intervention; (ii and iii) at "short-run" endline surveys 2 and 5 weeks after the grants; and (iv and v) at two "long-run" endline surveys 12 and 13 months after grants.[27] We ran pairs of surveys to reduce noise in outcomes with potentially low autocorrelation such as earnings or criminal activity. To measure time preferences, risk aversion, and baseline cognitive abilities (such as executive function), following each survey the respondents also conducted 45 minutes of incentivized games and tests (see Appendix D for measurement details). The winnings from all survey activities equalled about a half day's wages.

Response rates This sample was mobile and difficult to track. Roughly 40% changed locations between surveys, many changing sleeping places every few weeks or nights. Just 30% had mobile phones. Most went by several aliases, and may have been on the run.

To minimize attrition, we collected extensive contact information (all known addresses, plus at least five close contacts), and went to extreme effort to locate each person, wherever they had moved, averaging three to four days of searching per respondent per survey.

We collected data on 92.7% across all endline surveys. Attrition is relatively unsystematic: treatment arms had similar response rates (within 0.4% of the control group) while a test of joint significance of all baseline covariates yields $p = 0.328$.[28]

---

[26] In this simple case, there is no role for self control skills, θ, in aggression. This is a drawback of adapting the pecuniary crime model, since STYL explicitly teaches men skills to regulate their emotions in charged, automatic situations. In some sense, then, STYL may not only change the underlying value of σ (the extent of one's desire not to engage in criminal activity) but also one's ability to ensure that expressed actions conform to the underlying preferences rather than succumbing to immediate temptation or anger. This is functionally equivalent to predictions associated with a larger σ.

[27] The exception is the 100 men in the pilot, which had a single "short run" survey 3 weeks after grants. Actual survey times were, on average, 2.2, 5.7, 55.4 and 61.1 weeks after grants. Surveys were 90 minutes long and delivered verbally by enumerators in Liberian English on handheld computers.

[28] See Appendix A.3 for tracking techniques, response rates by survey wave and treatment group, and

Qualitative data We collected longitudinal qualitative data to better understand the context, intervention, and mechanisms. First, a Liberian research assistant acted as a participant-observer during the Phase 1 therapy. Second, we interviewed facilitators for their impressions of the intervention and participants. Third, three Liberian research assistants conducted semi-scripted interviews, 14 pre-treatment and 130 post-treatment, with 66 men in the sample.[29] Interviews covered job satisfaction, investments, economic challenges, plans, antisocial behaviors, and perceptions of the interventions.

5 Empirical strategy and estimation

We estimate intent-to-treat (ITT) effect on outcomes, Y, via the OLS regression:

$$Y_{ij} = \tau_1 TherapyOnly_i + \tau_2 CashOnly_i + \tau_3 Cash\&Therapy_i + \mathbf{X}_i\lambda + \gamma_j + \varepsilon_{ij} \qquad (1)$$

where $TherapyOnly$, $CashOnly$, and $Cash\&Therapy$ are indicators for random assignment to treatment arms: therapy only, cash only, or both therapy and cash. We control for a vector of baseline characteristics, X, and fixed effects for each of the j randomization blocks, γ_j. Y_{ij} is the average of the two proximate survey rounds (e.g. the 2- and 5-week surveys for short term effects). To reduce sensitivity to outliers, we top-code continuous variables at the 99th percentile. We test sensitivity to alternative approaches in Appendix E.1.

Key outcomes and multiple comparisons Our theory emphasizes five major outcome families: (1) anti-social behaviors, including crime and various forms of aggression; (2) income; (3) skills of self regulation and control; (4) economic time preferences; and (5) anti-criminal and antiviolent values. We identified the latter three as prime mechanisms. For each of the five outcomes we typically have multiple measures (e.g., earnings, assets, and consumption as measures of income). To reduce the number of hypothesis tests, we combine related measures into mean effects summary indexes, one for each major outcome.[30] We do

---

correlates of attrition. Of the 298 non-responses (of 3,896), we (i) had no location information (75%); men were mentally incapacitated (1%); died (8%, or 9 men); were in prison (12%); or refused (3%). Covariates associated with higher attrition include better mental health and income.

[29]19 in control, 16 in therapy, 15 in cash, and 16 in therapy then cash. Sampling was purposeful, based on variation in key baseline measures: economic success, crime, drug use, and present bias.

[30]We take averages of our outcome measures, coded to point in the same direction, akin to the approach by Kling et al. (2007). Note that the outcomes used to create the summary index may themselves be composites of many survey questions, such as consumption (a composite of many goods) or an aggressive behavior index (a composite of many types of aggressive behavior, a standard way that psychologists measure aggression). We do so because it is typically the composite itself rather than its component survey questions that we have theoretical interest or priors. In most cases this is reflected in the survey design, where the survey questions in each composite measure comprise a separate survey section. Also, to create an index by averaging the

15

not adjust p values for these five comparisons. The most conservative approach, a Bonferroni correction, would substitute .02 and .01 for the usual 0.1 and 0.05 critical values. In general, we will see the main findings pass even this conservative test.

This theory and the main outcomes were pre-specified but not formally registered. Our study predated the social science registry, but we outlined predictions for the main measures in a 2012 National Science Foundation proposal 1225697.[31]

Naturally these five are not the only conceivable outcomes of interest. There are alternative mechanisms (e.g. therapy could affect behavior by changing peers, post-traumatic stress, or substance abuse). Other outcomes enhance our interpretation (such as the effects of cash on economic inputs and outputs that generate income). We collected survey data for many such outcomes, and every one of these is reported in this paper's tables. But note that the treatment effects on these other outcomes are exploratory and suggestive, primarily meant for for interpretation and hypothesis generation.

Self-reported data   One threat to identification comes from systematic measurement error in self-reported data, especially measurement error correlated with treatment status. In the absence of administrative data such as arrest records, we developed a technique to validate select survey variables through intensive observation. Blattman et al. (2015) reports the approach in detail, and we summarize in Section 7 and Appendix F.

Spillovers   Another threat to identification comes from spillovers. Our recruiting strategy—working in large neighborhoods, recruiting less than 1% of adult men in those areas, and less than 15% of high-risk men we could identify on the street—was designed to reduce equilibrium effects such as a change in the returns to illicit work. We do not have the data or research design, however, to confirm that these equilibrium effects were minimized.

Another potential spillover involves interactions within and between treatment arms, especially therapy. For example, because of peer effects and the emphasis on social norms, there could be positive spillovers from treating groups of friends. If so, the coefficients on therapy would overestimate the effect of therapy in isolation. Alternatively, to the extent that control subjects interact with and learn from treatment subjects, they may acquire some of the lessons, leading us to underestimate therapy's impact.

---

component variables would give more weight to outcomes that are typically measured with many different questions (such as aggressive behavior) versus one that can be precisely measured with a small number of variables (such as drug selling), which we find inappropriate. Nonetheless, Appendix E.1 shows robustness to an index that averages all survey questions rather than composite measures, or uses covariance weighting rather than mean effects.

[31]http://chrisblattman.com/documents/research/2012.01.13_STYL_NSF_proposal.pdf. The fifth outcome, values, arose out of the review process and other comments and does not appear in this document.

We designed recruitment to minimize such interaction bias, but could not eliminate it. We do not have detailed social network data for the full sample, but we did trace social networks within the first two therapy groups. On average, each subject was acquainted with 6 of the 43 others in therapy, and 30% reported one close associate in therapy. Given randomization, we can assume similar relationships in the other arms. Without systematic data on networks we cannot estimate spillovers, and this is a weakness of our design. The two effects should cancel each other out somewhat, but the extent is unknown.

Interpretation and generalizability Another point is that our sample is not drawn from a well-defined population. This is a function of the setting—there is no administrative record of high-risk men in Liberia (or in any low-income or fragile state). We recruited men in a relatively transparent, replicable fashion, but a third declined to enter the study for reasons we cannot observe. Thus the treatment effects we estimate cannot be generalized to a defined population. This is not only a constraint of the setting, but also the nature of a proof-of-concept trial, where we have two promising but highly uncertain treatments— unconditional cash and CBT. Thus our study is akin to a medical efficacy trial, to determine whether the intervention produces the expected result under favorable circumstances.

6   Results

Figure 1 reports ITT estimates, from equation 1, of the effect of each treatment on the five main summary indexes. We discuss each in turn.[32]

6.1   Antisocial behaviors

Table 3 reports program impacts on self-reported anti-social behaviors. We defined these as disruptive or harmful acts towards others, such as crime or aggression. The family excludes self-harm (e.g. drug abuse) or the acts of peers, outcomes, which we analyze in section 6.6.

Cash did not lead to a statistically significant or sustained reduction in antisocial behaviors, but therapy did. Therapy led to large reductions in the short run, by 0.25 standard deviations with therapy alone and 0.31 standard deviations with therapy plus cash. This reduction in antisocial behaviors persisted, however, only when therapy was followed by cash:

---

[32]As seen in Appendix E.1, these effects are highly robust to a variety of specifications and attrition scenarios. We obtain similar results if we: pool the endlines rather than averaging them; construct summary indexes of all underlying survey questions rather than indexes of the composite measures; or covariance weight rather than weight index components equally. We also show that the results are robust to conservative attrition scenarios by substituting extreme values for missing outcomes.

Figure 1: Program impacts with 95% confidence intervals (summary indexes, z-scores)
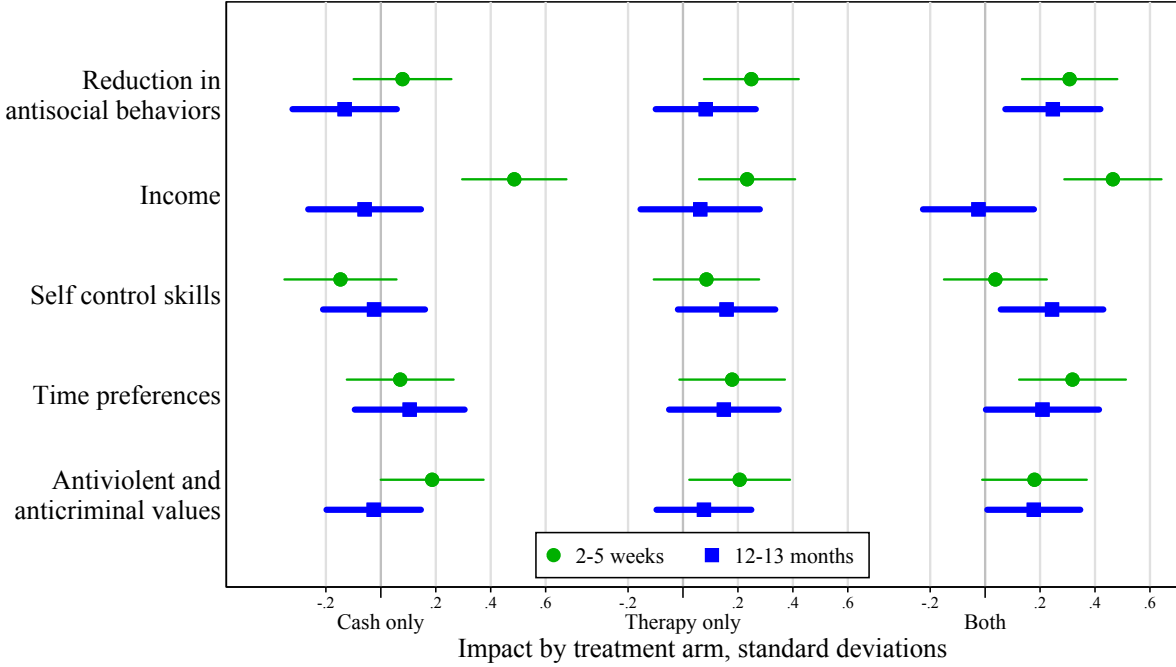
Notes: The figure reports the effect of each treatment arm in the short run (2 & 5 weeks) and long run (12 & 13 months). Treatment effects are estimated via OLS controlling for baseline covariates and block fixed effects. Each summary index is the standardized mean of composite outcomes, where we have changed the sign of all indexes to point in a positive direction. Standard errors are heteroskedastic-robust. *** p<0.01, ** p<0.05, * p<0.1

one year after therapy, therapy alone led to a 0.08 standard deviation fall in antisocial behaviors (not statistically significant) compared to a 0.25 standard deviation fall with therapy plus cash (significant at the 1% level). This difference between therapy and therapy plus cash in the long term is significant at the 5% level.[33]

Turning to the component measures, we must interpret individual estimates with caution. Nonetheless the coefficients on therapy only or therapy plus cash are almost universally negative, and a majority are statistically significant.

Drug selling and other crime. In the short run, 17% of the control group said they sold drugs, and they admitted to 2.6 acts of theft or robbery in the previous two weeks. In the long run survey, 13.5% sold drugs and they reported 1.9 acts of theft. Crime rates could fall in the control group to the extent that we are recruiting people in especially hard times (i.e.

---

[33]See Appendix E.2 for formal tests of the difference between both therapy and cash to therapy or cash alone. Appendix E.3 tests whether short and long run impacts are equal. We cannot reject the hypothesis that the short and long run effects of both therapy and cash are equal over time, but we can reject the equality of effects for therapy alone.

Table 3: Program impacts on antisocial behaviors

Outcome	Round	Control mean	ITT regression (N= 947)					
			Therapy only		Cash only		Both	
			ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Antisocial behaviors, z-score	2–5w	0.151	-0.249	[.088]***	-0.079	[.091]	-0.308	[.089]***
	12–13m	0.032	-0.083	[.093]	0.132	[.097]	-0.247	[.088]***
Usually sells drugs	2–5w	0.166	-0.077	[.027]***	-0.041	[.029]	-0.076	[.028]***
	12–13m	0.135	-0.034	[.029]	0.035	[.030]	-0.059	[.029]**
# of thefts/robberies in past 2 weeks (sum of 8)	2–5w	2.577	-0.841	[.400]**	-0.770	[.409]*	-1.236	[.407]***
	12–13m	1.839	0.073	[.395]	0.352	[.388]	-0.728	[.363]**
Disputes and fights in past 2 weeks (mean of 9), z-score	2–5w	0.076	0.013	[.092]	0.027	[.091]	-0.132	[.076]*
	12–13m	-0.060	-0.026	[.091]	0.100	[.090]	-0.100	[.077]
Carries a weapon on body[†]	2–5w	0.157	-0.086	[.034]**	-0.045	[.037]	-0.093	[.035]***
	12–13m	0.148	-0.059	[.031]*	0.043	[.035]	-0.066	[.033]**
Arrested in past 2 weeks	2–5w	0.139	-0.011	[.027]	0.006	[.027]	-0.013	[.029]
	12–13m	0.118	-0.006	[.024]	0.007	[.025]	-0.033	[.024]
Aggressive behaviors (mean of 19), z-score	2–5w	0.102	-0.208	[.081]**	0.008	[.085]	-0.196	[.087]**
	12–13m	0.188	-0.153	[.110]	-0.043	[.107]	-0.339	[.109]***
Verbal/physical abuse of partner (mean of 4), z-score[†]	2–5w	-0.035	-0.087	[.111]	0.091	[.114]	-0.032	[.115]
	12–13m	-0.071	0.142	[.100]	0.233	[.113]**	0.059	[.104]

Notes: The table reports intent to treat estimates of the effect of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. For instance, thefts/robberies is the sum of 8 kinds of crimes; disputes/fights is the standardized mean of 9 kinds of physical or verbal altercations with peers, community, and authorities; aggressive behaviors is the standardized mean of 19 possible types of aggression and hostility; and verbal and physical abuse of partners is the standardized mean of 3 forms of verbal abuse of intimate partners plus one form of physical abuse. (For the latter two cases, we report standardized indexes since the incidents are measured on a 0–3 frequency scale, and the absolute sum itself has no interpretation.) The overall summary index is the standardized mean of these seven composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

[†]These variables were not collected during every phase/round, so their regressions have a smaller sample size.

there is regression to the mean). With therapy, however, crime rates fell by roughly 40% in the short run, and this fall persisted in the long run with therapy plus cash. Appendix D elaborates on individual crime measures. If we extrapolate these results to the full year since baseline, therapy plus cash led men to go from 61 to 30 drug sales and robberies per person per year (see Appendix E.5). Given the $530 cost of the two interventions, this is roughly $21 per crime in the first year alone, ignoring any ongoing impact on crime, or impacts of the program on aggression, arrests, or incomes.

Fights We also asked about 9 types of verbal and physical altercations in the past two weeks, including the frequency and severity of disputes with peers, neighbors, community leaders, or the police. Here, as with all summary indexes in the paper, we use the standardized mean effects of all nine survey questions.[34] On average, the men reported about one dispute in the past two weeks. The decline from therapy alone or therapy and cash is not statistically significant, though the point estimate on therapy and cash is negative in the long run: -0.10 standard deviations.

Weapons We asked men if they carried a weapon on their body for protection. This was typically a knife, as guns were rare. In the long run, 15% were carrying a weapon, and this fell about 6 percentage points with either therapy alone or therapy plus cash.

Arrests 12% of the control group reported an arrest in the two weeks prior to the endline survey. We did not see a statistically significant decline in arrests, though after one year the coefficient on therapy plus cash represented a 28% decline, of about one arrest per year.

Aggressive and hostile behaviors We asked 9 questions from a scale of reactive and proactive aggression (adapted to Liberian English by the authors) such as the frequency with which they yell, curse, or lose tempers (Raine et al., 2006). Based on our qualitative work, we added 10 more locally-relevant acts, such as cheating someone, threatening others, or bullying. In the long run, the index of all 19 questions fell .15 standard deviations (not significant) with therapy alone and .34 with both (significant at the 1% level).

Intimate partner abuse We collected a crude measure of intimate partner abuse—3 questions on verbal abuse (such as cursing and yelling) and one on incidents of physical abuse in the past two weeks. A standardized index of these measures fell little in the short

---

[34]A main reason is because the measurement scales differ across component survey variables and the absolute valuer of the scales themselves are not meaningful (e.g. a frequency scale of 0–3, from never to often) We standardize individual survey questions, average them, and standardize this composite to have mean zero and unit standard deviation. Results are robust to alternate weighting and indexing approaches.

run with therapy, and in the long run the coefficient on therapy plus cash is actually positive (the only instance where therapy is positively correlated with violence).

Political violence   Given Monrovia's recent history of mercenary recruitment, minor riots, and some election violence, we predicted the men would have opportunities for such violence. Indeed, shortly after the small group of Phase 1 men received therapy, there was a minor riot in the city.[35] From then, however, Liberia entered one of the most politically quiescent periods in recent history, and so we effectively have no incidents of political violence to measure. This is the only pre-specified outcome that we could not test directly.

Heterogeneity   A natural question is whether the therapy is impactful for the most or least antisocial men. Appendix E.6 reports ITT regressions where we add an interaction between the treatment indicators and a standardized index for antisocial behavior at baseline. The therapy was impactful for the average participant, but the greatest decline in antisocial behavior was among those with the highest initial antisocial behaviors.

6.2   Income

Table 4 reports program impacts on income as well as related economic activities, to aid interpretation. We measured income in three ways: (i) estimated earnings in all activities in the past two weeks; (ii) consumption in the past two weeks; and (iii) an index of durable assets.[36] A summary index of all three measures rises by .49 standard deviations in the short run from cash alone and .47 standard deviations with cash and therapy. But after a year there was no significant difference in income from any treatment.[37]

To explore why incomes rose then fell, we turn to other economic outcomes. Consumption and assets could rise simply from spending the grant. But this doesn't explain the rise in short run earnings—by as much as a third in the short run from cash alone. Overall, the cash seems to have been saved and invested in petty business, and this accounts for the rise in short run earnings. But bad shocks, especially theft, meant these gains were fleeting.

---

[35]The men in all three treatment arms were slightly less likely to participate or sympathize with the rioters, but with a sample size of just 100 these effects were not significant.

[36]All measures were pre-specified. To obtain earnings, we first asked each respondent their gross and net earnings in the past four weeks across 25 economic activities (legal and illegal). This earnings measure could still be subject to recall and other biases, and may inadequately capture home production. Thus we also use two measure of permanent income. One is an index of durable assets—a z-score constructed by taking the first principal component of 42 measures of land, housing quality, and small and large household assets. We also conduct an abbreviated consumption module of short-term food and non-food consumption.

[37]Other outcomes are consistent with these patterns. Homelessness falls in the short run as income rises, but there is no long term effect. Savings also jumps substantially in the cash only and therapy plus cash groups. Unlike income, however, this savings impact persisted in the long run in the therapy plus cash arm,

Table 4: Program impacts on income and economic activity

Outcome	Round	Control mean	Therapy only		Cash only		Both	
			ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
						ITT regression (N=947)		
Summary index of income measures, z-score†	2-5w	-0.240	0.233	[.089]***	0.486	[.097]***	0.465	[.090]***
	12-13m	0.031	0.063	[.111]	-0.059	[.105]	-0.024	[.103]
Weekly earnings in past two weeks (USD)	2-5w	14.410	1.351	[1.592]	4.128	[1.572]***	3.036	[1.535]**
	12-13m	17.595	0.764	[1.723]	1.341	[1.663]	0.436	[1.882]
Durable consumption assets, z-score†	2-5w	-0.093	0.124	[.094]	0.138	[.094]	0.240	[.097]**
	12-13m	0.009	0.143	[.096]	-0.111	[.097]	0.086	[.096]
Consumption, past 2 weeks, USD	2-5w	44.394	10.683	[3.812]***	24.040	[4.199]***	20.695	[3.553]***
	12-13m	47.347	-1.870	[3.808]	-2.615	[3.605]	-5.005	[3.589]
OTHER ECONOMIC OUTCOMES:								
Savings stock (USD)†	2-5w	45.957	-1.652	[8.901]	16.342	[10.018]	18.131	[10.169]*
	12-13m	51.395	11.360	[11.042]	2.026	[10.245]	21.744	[11.352]*
Homeless in past two weeks	2-5w	0.202	0.004	[.029]	-0.090	[.029]***	-0.100	[.029]***
	12-13m	0.145	0.018	[.030]	0.016	[.029]	-0.020	[.030]
Investment in past two weeks (USD)	2-5w	18.444	6.748	[6.291]	55.750	[8.410]***	48.122	[7.797]***
Value of business assets (USD)†	12-13m	26.121	3.116	[13.431]	19.391	[15.276]	13.548	[13.007]
Hours/week of work, past month	2-5w	36.773	1.044	[2.879]	7.439	[2.907]**	1.787	[2.966]
	12-13m	34.273	1.030	[2.550]	0.681	[2.525]	-1.416	[2.493]
Home robbed or belongings stolen in past month†	2-5w	0.859	-0.011	[.035]	-0.009	[.036]	-0.012	[.038]
	12-13m	0.797	-0.037	[.043]	-0.086	[.042]**	0.013	[.042]

Notes: The table reports intent to treat estimates of the effect of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. The income summary index is the standardized mean of three composite outcomes (themselves first standardized). Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

†These variables were not collected during every phase/round, so their regressions have a smaller sample size.

Table 5: Self-reported allocation of the grant, by expenditure category

Expenditure category	Treatment arm		Difference (N=475)	
	Cash & therapy	Cash only	Coeff.	p value
	(1)	(2)	(3)	(4)
Living expenses (such as food, clothing, rent)	28%	26%	0.04	0.13
Durable assets	7%	6%	-0.01	0.53
Drugs, alcohol, gambling & sex	4%	4%	0.00	0.95
Gifts and transfers to others	11%	11%	0.01	0.65
Business investments and expenses	23%	24%	-0.04	0.20
Savings and debt payments	20%	20%	0.02	0.57
Own health and education	8%	8%	0.00	0.89

Notes: Using pictures of different types of spending and plastic chips, grant recipients indicated how they used the grant. Columns (1)–(2) report the means for each treatment arm. Columns (3)-(4) report the coefficients and p values from an OLS regression of the proportion spent on an indicator for assignment to therapy then cash controlling for block fixed effects and baseline covariates.

We assessed grant spending in two ways. Using pictures of different types of spending and plastic chips, we first asked grant recipients to indicate how they used the grant. Table 5 reports self-reported allocations of the grant by treatment arm. We see little effect of the recent therapy on allocations. Little of the grant seems to have spent on drugs, alcohol, gambling and prostitution. Even if men underreport these expenses, we see no difference between cash recipients who did and did not receive therapy.

We can also look at expenditure data, which included a range of business investments in the two weeks prior to the 2- and 5-week surveys. As reported in Table 4, those who received only cash reported $56 more investment in each 2-week period. Thus the total 5-week investment treatment effect is at least $112—around 60% of the grant. Meanwhile, the therapy only group resembled the control group in terms of investment. Labor inputs also rise, as total hours of work per week rise by 20% in the short run.

These short run investments did not last. In the cash only group, the stock of business assets after a year is only $19 greater than in the control group, not statistically significant. We also see no long run difference in total work hours.[38]

What happened? From qualitative interviews, insecure property rights were a major barrier to capital accumulation. A large number of men reported the theft of all their assets, or all their wares, on a regular basis, by criminals or (for market wares) the police.[39] At each

---

rising about 36% relative to the control mean.

[38] All three treatments caused individuals to substitute from illicit work to non-agricultural low-skill business in the short term, but the effects were most pronounced and longer lasting for the group that received both cash and therapy (see Appendix E.7).

[39] In some cases this was theft by a friend, family member, or stranger. Also common was confiscation of

survey round, about 70% of the men reported a house robbery and belongings stolen in the past month. This implies a robbery every other month, at least. There is little difference by treatment status, suggesting that men were not more likely to be targeted if they received cash. But they would have had more to lose.

6.3 Self control skills

Table 6 examines our proxies for the three mechanisms outlined in our model: self control, time preferences, and anticriminal/antiviolent values. We measured self control skills using standard psychometric questionnaires for four constructs that psychologists associate with less impulsive and more planful behavior.[40]

We see long term improvements in self control of about the same magnitude as the reduction in antisocial behavior. After one year, an overall summary index of these skills increases 0.16 standard deviations with therapy only (significant at the 90% level) and 0.24 with therapy plus cash (significant at the 95% level). The short-term changes are positive but, strangely, are not nearly as large or robust as the long term changes (though large long term changes are within the confidence interval).

This summary measure has four components. First, we looked at 9 questions from the Barrett Impulsiveness Scale (Spinella, 2007), which assesses one's inability to control thoughts and actions.[41] Second, we used 8 questions from the NEO-five factor personality inventory to assess conscientiousness (Costa and McCrae, 1997). Topics included following societal rules, and controlled, careful behavior. Third, we took 7 questions on perseverance from the GRIT scale (Duckworth and Quinn, 2009), which captures the ability to press on in the face of difficulty. Finally, we selected 8 questions on reward responsiveness—whether they are motivated by immediate, typically emotional rewards—from the Behavioral Inhibition/Behavioral Activation Scale.[42] Appendix D lists all questions.

Therapy led to long term reductions in both impulsivity and reward responsiveness. Conscientiousness and GRIT also improve, but the magnitudes are more modest and the results are not significant. While we do not want to over-interpret the index components, the pattern is consistent with STYL having a greater effect on immediate self-control in the

---

wares by the police. Some forms of market selling contravene official rules, often unenforced, but nonetheless giving police opportunities to confiscate. Some confiscation is legitimate, some not.

[40]We translated the instruments into Liberian English, pretested them outside our sample, and then adapted the questions to the context or dropped inappropriate ones (as the standard questionnaires typically offered dozens of possible questions).

[41]Examples include "I buy things on impulse" or "I say things without thinking".

[42]Examples include "I will often do things for no other reason than that they might be fun" or "When I see an opportunity for something I like I get excited right away.". Previous research has linked disruptions in and extremes of reward motivation to substance abuse (Robinson and Berridge, 2000).

Table 6: Program impacts on self control skills, time preferences, and values

Outcome, z-score	Round	Control mean	Therapy only ITT	Therapy only Std. Err.	Cash only ITT	Cash only Std. Err.	Both ITT	Both Std. Err.
			ITT regression (N= 947)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Self-control skills $(\theta)^{\dagger}$	2–5w	-0.037	0.085	[.098]	-0.147	[.104]	0.037	[.096]
	12–13m	-0.070	0.159	[.090]*	-0.025	[.095]	0.244	[.095]**
Impulsiveness	2–5w	-0.010	-0.011	[.101]	0.180	[.108]*	0.104	[.095]
	12–13m	0.082	-0.178	[.096]*	0.006	[.098]	-0.212	[.099]**
Conscientiousness	2–5w	-0.077	0.109	[.105]	0.046	[.106]	0.163	[.105]
	12–13m	0.018	-0.065	[.097]	-0.028	[.100]	0.044	[.097]
Perseverance/GRIT	2–5w	-0.035	0.027	[.099]	-0.130	[.105]	0.042	[.104]
	12–13m	-0.037	0.116	[.099]	0.057	[.099]	0.105	[.103]
Reward responsiveness	2–5w	-0.010	-0.071	[.106]	0.107	[.107]	0.013	[.105]
	12–13m	0.072	-0.165	[.102]	0.084	[.100]	-0.242	[.102]**
Forward-looking time preferences (δ, β)	2–5w	-0.202	0.179	[.098]*	0.071	[.099]	0.318	[.099]***
	12–13m	-0.149	0.149	[.102]	0.105	[.102]	0.209	[.105]**
Patience	2–5w	-0.093	0.187	[.073]**	0.116	[.073]	0.267	[.074]***
	12–13m	-0.240	0.170	[.103]*	0.145	[.096]	0.258	[.099]***
Time inconsistency	2–5w	0.008	-0.063	[.074]	-0.009	[.076]	-0.138	[.075]*
	12–13m	0.129	-0.072	[.083]	0.018	[.087]	-0.059	[.084]
Anticriminal & antiviolent values $(\sigma)^{\dagger}$	2–5w	0.100	-0.206	[.094]**	-0.187	[.096]*	-0.180	[.097]*
	12–13m	0.070	-0.076	[.088]	0.026	[.088]	-0.177	[.086]**
Attitudes toward use of violence	2–5w	0.021	-0.141	[.105]	-0.201	[.106]*	-0.057	[.107]
	12–13m	0.051	0.019	[.108]	0.080	[.109]	-0.046	[.109]
Attitudes toward criminality	2–5w	0.139	-0.177	[.107]*	-0.154	[.112]	-0.242	[.110]**
	12–13m	0.044	-0.062	[.102]	-0.041	[.100]	-0.244	[.101]**
Attitudes toward political violence	2–5w	0.111	-0.224	[.109]**	-0.136	[.108]	-0.173	[.113]
	12–13m	0.096	-0.119	[.105]	0.012	[.105]	-0.167	[.106]

Notes: The table reports intent to treat estimates of the effect of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Program impacts on noncognitive skills and values according to their coverage in the STYL curriculum, 12–13 month only

Outcome (# of question in index), z-score	Control mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
		ITT regression (N=947 subjects)					
		Therapy only		Cash only		Both	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of self-control skills							
Topics emphasized in curriculum (16)	-0.092	0.169	[.090]*	0.026	[.091]	0.170	[.093]*
Topics not emphasized in curriculum (16)	-0.024	0.054	[.098]	-0.070	[.101]	0.232	[.099]**
Summary index of anticriminal/antiviolent values[†]							
Topics emphasized in curriculum (8)	0.041	-0.085	[.107]	0.012	[.110]	-0.116	[.108]
Topics not emphasized in curriculum (21)	0.070	-0.062	[.087]	0.028	[.087]	-0.184	[.086]**

Notes: The table reports intent to treat estimates of the effect of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We have subdivided the summary indexes reported in Table 6 by their coverage of the specific topics in the STYL curriculum. *** p<0.01, ** p<0.05, * p<0.1

[†]These variables were not collected during every phase/round, so their regressions have a smaller sample size.

moment rather than self-discipline over longer term goals. Many psychologists also think of conscientiousness and GRIT as stable traits that are unlikely to change after adolescence, and the results are somewhat consistent with that view. The short-term impacts are positive but, strangely, are not nearly as large or robust as the long term impacts. We don't know why, though the large long term changes are within the confidence interval of the short term effect. It could be that impacts on self control solidify over time.

We must be cautious because these scales are self-reported, and treated men could simply be repeating back their lessons. There is some evidence this is not so. We divide the 32 self control questions into two indexes: questions with high (44%) and low (56%) emphasis in the curriculum.[43] Table 7 reports the ITT estimates after a year. The effect of cash and therapy is at least as large for low emphasis items.

6.4 Time preferences

We measured the degree of forward-looking time preferences via both incentivized games and survey questions, and we report a summary index of 4 measures of patience and 4 of time inconsistency—akin to δ and β in our model. Initially, we report these separately from

---

[43]We rated each index component on a scale of 0 (not emphasized) to 4 (very emphasized). We then defined low-emphasis components as those rated 0 or 1 and high emphasis components as those rated 2 or above. These results are unchanged for using 1.5 or 2 as the emphasis cutoff.

"self control" for fidelity to the model, and because they are distinct in measurement.

Specifically, we have: (i) a set of incentivized tradeoffs between modest amounts of money now versus in two weeks, and again in two versus four weeks, that allow us to place men in seven ordered bins of patience and time-inconsistency (for an average payout of $3, about a day's wages); (ii) a hypothetical (non-incentivized) version of the same tradeoffs, with higher stakes tradeoffs; and (iii) self-reported assessments of time preferences. All are described in Appendix D.3. In the spirit of reporting and analyzing all survey measures, the summary index in the main table includes all three types of measures. The family index, which we call forward-looking time preferences, includes both patience and time-inconsistency.[44]

In the short term, the time preferences index increases for all treatment arms, though the result is smallest and the least statistically significant for cash only (possibly a liquidity effect on game play). In the long run, those who received therapy alone are 0.15 standard deviations more forward-looking, and those who received both are 0.21 standard deviations more forward-looking. These results seem to be strongest among the set of incentivized and hypothetical measures of patience rather than time inconsistency (see Appendix D.3).

Time preferences enter into the theoretical model differently than self control, but an obvious question is whether they are distinct. The correlation between the self control and time preference summary indices is 0.33, and is significant at the 1% level. If we combine the time preference and self control measures into a single summary index, therapy alone and therapy plus cash both have highly statistically significant positive impacts of roughy 0.2 standard deviations (Appendix D.6).

Finally, if we consider the incentivized measures of time preference alone, the treatment effects are in the same direction but lower in magnitude, and not statistically significant (see Appendix D.3). Hence we regard the overall index results with some caution.

6.5   Anti-criminal and anti-violent values

For values, we measured self-reported attitudes towards crime and violence in the men's own lives—indicators of the degree to which they had internalized mainstream social norms.[45] We asked 11 questions on attitudes to the use of violence to solve community or personal problems, such as mob killings of suspected thieves, or attacking their unfaithful wife's lover. We also asked 12 questions about their attitude toward participating in crime, including whether they would feel fine taking unwatched goods or stealing $100 from someone's pocket.

---

[44]We do so to reduce hypothesis tests, because they are conceptually related, have similar comparative statics, and self-reported assessments are not cleanly divided between patience and time-inconsistency.

[45]For anticriminal/antiviolent values, we also have proxies for changes in self-image, but since self-image itself was not among the theoretical or empirical measures we pre-specified, we consider it alongside other outcomes in the next section.

We also asked about 6 hypothetical forms of political violence, including whether they discuss protesting with friends or making trouble or conflict with the authorities (See Appendix D). We did not measure perceived social category directly.

An index of all three composites shows that all treatments decreased the self-reported acceptability of violent behaviors in the short run. Cash had little impact on self-reported attitudes in the long run, but therapy plus cash led to a .18 standard deviation decline. We cannot reject that the effects of therapy only and therapy plus cash are equal. The overall effect is driven by attitudes toward criminality and political violence.

As with self control, we divide the 29 questions into two indexes by high and low emphasis in the curriculum. Table 7 reports the ITT estimates after a year. The effect of cash and therapy is at least as low for the low emphasis components.

6.6   Other outcomes of interest

NEPI's therapy targeted other behaviors, especially ones they thought could contribute to antisocial behavior, including substance abuse or harmful peers. These possible mechanisms and also important outcomes. Table 8 reports impacts on all these other measures.

Prosocial behavior   We measured various prosocial behaviors in the long run surveys (including group memberships, group and community leadership, contributions to local public goods, and trust in others). In contrast to the steep reductions in antisocial behaviors, we see no evidence that therapy or cash led to more positive social engagement.

Mental health   We asked about two disorders we deemed relevant. We measured five symptoms of post-traumatic stress using existing Liberian instruments.[46] We also measured neuroticism, the tendency to experience emotional instability or anxiety, assessed with 8 questions from the NEO-5 factor personality inventory (Costa and McCrae, 1997). After one year, the therapy plus cash group reported 0.17 standard deviations lower post traumatic stress and 0.15 standard deviations lower neuroticism (though both are not significant).

Self efficacy and esteem   The therapy did not treat these traits directly, and we have no theory to suppose a direct effect, but both could improve with self control and image change, and both could reinforce reductions in antisocial behaviors.[47]   We asked eight questions from a standard locus of control questionnaire, which aims to measure the extent to which

---

[46]We used the five symptoms with the highest factor loadings in surveys of ex-combatant mental health by Blattman and Annan (2015).

[47]Negative self image has been linked with many aspects of negative behavior and counterproductive or extreme risk-seeking behavior (Coopersmith, 1967).

Table 8: Impacts on other outcomes

Outcome, z-score	Round	Control mean	Therapy only ITT	Therapy only Std. Err.	Cash only ITT	Cash only Std. Err.	Both ITT	Both Std. Err.
				ITT regression (N=947)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prosocial behavior	12–13m	0.018	0.041	[.088]	-0.075	[.085]	-0.017	[.090]
Post-traumatic stress (5 questions)[†]	2–5w	0.067	-0.031	[.096]	-0.013	[.101]	-0.127	[.101]
	12–13m	0.136	-0.124	[.101]	-0.061	[.100]	-0.167	[.104]
Neuroticism (8 questions)[†]	2–5w	0.057	0.011	[.098]	0.021	[.101]	-0.077	[.104]
	12–13m	-0.019	0.044	[.097]	0.035	[.102]	-0.153	[.096]
Locus of control (8 questions)[†]	2–5w	-0.001	0.000	[.098]	0.057	[.104]	-0.089	[.100]
	12–13m	0.010	-0.032	[.101]	-0.111	[.098]	-0.022	[.106]
Self esteem (8 questions)[†]	2–5w	-0.099	0.112	[.099]	-0.012	[.100]	0.191	[.102]*
	12–13m	-0.071	0.078	[.098]	0.060	[.100]	0.190	[.101]*
Appearance (6 questions)	2–5w	-0.118	0.085	[.081]	0.131	[.081]	0.203	[.080]**
	12–13m	0.016	-0.102	[.078]	-0.085	[.077]	-0.109	[.082]
Substance abuse (0-3 index)	2–5w	1.378	-0.249	[.070]***	-0.060	[.072]	-0.197	[.071]***
	12–13m	1.091	-0.065	[.061]	0.063	[.062]	-0.057	[.060]
Quality of social networks	2–5w	-0.241	0.147	[.062]**	0.109	[.061]*	0.325	[.062]***
	12–13m	0.066	0.063	[.092]	-0.044	[.092]	0.139	[.095]
Peers (20 questions)[†]	2–5w	-0.160	0.207	[.091]**	0.014	[.095]	0.235	[.094]**
	12–13m	0.040	0.011	[.088]	-0.070	[.089]	0.017	[.090]
Family (4 questions)[†]	2–5w	-0.192	0.106	[.099]	0.131	[.105]	0.307	[.099]***
	12–13m	-0.019	0.124	[.099]	0.070	[.100]	0.129	[.097]
Ex-commanders (4 questions)	2–5w	-0.141	-0.011	[.038]	0.013	[.038]	-0.038	[.038]
	12–13m	0.176	0.004	[.076]	0.026	[.078]	-0.139	[.074]*
"Big men" (5 questions)[†]	2–5w	-0.012	0.015	[.104]	0.039	[.107]	0.172	[.107]
	12–13m	0.120	0.001	[.155]	-0.130	[.152]	0.071	[.160]
Subjective well being (3 questions)	2–5w	-0.237	0.166	[.087]*	0.170	[.087]*	0.425	[.094]***
	12–13m	-0.020	0.057	[.072]	-0.009	[.072]	0.184	[.074]**
Executive function	2–5w	-0.103	0.076	[.075]	0.059	[.077]	0.024	[.085]
	12–13m	0.110	-0.094	[.077]	-0.078	[.076]	-0.109	[.078]

Notes: The table reports intent to treat estimates of outcomes that were not a priori specified as of primary interest. We calculate the impact of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heteroskedastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

[†] These variables were not collected during every phase/round, so their regressions have a smaller sample size.

individuals believe they can control events affecting them (Sapp and Harrod, 1993). We also asked eight questions about self esteem, such as, "I am able to do things as well as most other people" or "I take a positive attitude toward myself." Therapy seems to have a positive effect on self esteem, and when followed by cash this change is large, 0.19 standard deviations, and persistent. We see no such change in locus of control, however.

Appearance The therapy encouraged men to change their appearance. At the end of the survey enumerators recorded their subjective impressions: quality of dress, shoes, cleanliness, and smell. We see a short run effect from therapy, but this is not sustained in the long run.

Substance abuse Therapy tried to equip participants with strategies to cut back substance abuse (but discouraged quitting "cold turkey" to reduce risk of withdrawal problems). At the one year endline, reports of daily use of in the control group are 76% for alcohol, 50% for marijuana, and 20% for hard drugs. An index of all three indicators (0–3) fell 0.20 in the short run but only 0.06 in the long run (not statistically significant).

Quality of social networks We also assessed risky social networks.[48] We measured the traits, positive and negative, of men's five closest peers.[49] We also asked about closeness to and support received from family members, former rebel commanders, and "big men" (intended to connote a criminal boss). A summary index of positive social networks increases in the short run by 0.15 standard deviation with therapy and 0.33 standard deviations with therapy plus cash. But these changes are no longer as large or significant a year later.

Subjective well being We asked men to rank their subjective well being in absolute terms and relative to others in their community.[50] All therapies show a positive short term effect but there is no long term effect of therapy only or cash only. Those who received therapy plus cash report 0.15 standard deviations greater subjective well being.

---

[48]In some settings, neighborhood changes would also indicate a change in risky behavior, but not in Liberia. Most high-crime neighborhoods in Monrovia are mixed-income residential and market areas where high-risk men are a small minority. They live on the margins, often in abandoned areas within these neighborhoods (building sites, forested groves, garbage dumps, and abandoned buildings). Men who turn their lives around do not need to move neighborhoods, but rather stay where they are but move in different, more mainstream social circles, avoiding high-risk hangouts. Thus we do not report on neighborhood movements.

[49]We ask men who their five closest peers are, by name, and then ask whether they hold any of 20 qualities ranging from positive (they work hard, save, go to school) or negative (the steal, do drugs, get in fights).

[50]We asked about well being, health, wealth, and power in absolute terms. We asked about wealth, respect, power, and access to services in relative terms. Each used a picture of a ladder with 10 rungs. The summary index is the average of each ladder. Patterns are broadly similar across all ladders.

Executive function   We developed tests of executive function—cognitive processes associated with inhibitory control, working memory, self regulation, and planning. Tests included a modified Simon task and digit recall (see Appendix D.5). We did not hypothesize a change because these are thought to be abilities that solidify in early childhood (Appendix D.6). As expected, there is no statistically significant change from treatment.[51]

6.7 Insights from qualitative interviews and observation

One of the strongest impressions we gained from qualitative interviews was the importance men attached to identity change, what they and NEPI called "transformation". Nearly all the subjects we interviewed described feeling ostracized at baseline, and many reported that the therapy pushed them to believe they could be someone better for the first time.

The facilitators played an important role here. The participants we interviewed unanimously had admiration and praise for the facilitators, highlighting that their backgrounds demanded respect and credibility among respondents, while their personal stories of change were encouraging. Beyond modeling the change in self image and social category, men reported the facilitators were also sometimes the first people to treat them with seriousness and respect, and this built their confidence to reintroduce themselves to community members, or to expose themselves to banks and shops.

Attempts to behave normally, especially the exposure to new social situations, seemed to reinforce skill and identity change. Many of the men failed in their plans, or experienced stigma in their shop or bank visits. In group sessions, men discussed what went wrong and why (such as poor decisions, or choice of dress). Men with setbacks learned from and were encouraged by the positive experiences of others. And facilitators sometimes observed men's homework attempts and coached them through difficulties.

Men's appearance also transformed during therapy. The first day men arrived with long or messy hair, facial hair, dirty or ripped clothing, wearing t-shirts with shorts and sandals. Their demeanor was tough, and their appearance signaled outcast status. Haircuts were offered in week two, and many men took advantage, symbolizing the change. Others showed up beforehand having gotten a haircut on their own. Similarly, before the unit on hygiene, some men began arriving in pants, shoes, and collared shirts. Typically a few men in each group resisted these changes. But seeing the positive experiences of others, they too began to arrive more clean cut, trying out the new image. The survey results confirm a short-term change in appearance. The absence of long term change is puzzling.

---

[51]We primarily measured executive function as a baseline control, and measured it over time for other research purposes. If we combine executive function, time preferences, self control skills into a single index, we still see sizable long term treatment effect of therapy and cash (Appendix D.6).

A year later, therapy participants also described applying skills of self-regulation in their lives. To avoid fights, they used new tactics: removing themselves from emotionally-charged situations, allowing space to process their feelings, and ignoring negative automatic thoughts in the favor of more controlled thinking. Related were improved social and communication skills. Interviewees described how such skills allowed them to engage with community members or in disputes and express themselves without anger or violence.

Not only did the community regard them differently, many said, but troubled young men began coming to them for advice and lessons learned from the therapy once they saw the sudden and sustained change–another important source of reinforcement, and perhaps one reason we do not see a change in peer quality in the data.

7 Can we believe our self-reported data?

Self-reported data raise several worries, the most serious being measurement error correlated with treatment. For instance, men who receive an anti-violence intervention might be more likely to tell us they are non-violent, overestimating the estimated treatment effect of therapy.

This kind of bias is hard to square with the patterns of effects we observed. Therapy followed by cash would have to induce systematic errors where therapy or cash alone did not. Nonetheless, this is possible. Thus, concerned that our survey measure, y^s, may be biased, we set out to intensively validate some measures, y^v. If y^v is closer to the true behavior, y^*, this allows us to estimate the degree and direction of bias. We summarize the approach, empirical strategy, and results here, with details in Appendix F and Blattman et al. (2015).

7.1 Approach to validation

Of more than 4,000 endline surveys, we randomly selected 7.3% and validated answers to six survey-based measures with two-week recall periods. We chose four potentially sensitive behaviors—marijuana use, thievery, gambling, and homelessness. We also chose two every-day expenditures that we did not consider sensitive but could be subject to recall bias or other error—paying to watch television in a video club, and paying to charge a mobile phone.

We used intense qualitative work—in-depth participant observation, open-ended questioning, and efforts to build relationships and trust—to try to elicit more truthful answers. Over several days of trust-building and conversation, plus direct observation, we tried to elicit a direct admission or discussion of the behavior.

We selected and trained eight of the study's most talented qualitative research staff as validators, all Liberians. In the ten days following the survey, a validator visited the respondent over four days, spending several hours each day in conversation and observation.

Validators shadowed respondents as they went about their day, rather than conduct formal interviews. They raised target topics through indirect questions while chatting.

Validators developed techniques to foster trusting relationships and to build rapport: becoming close to street leaders; eating meals with subjects; sharing personal information (including similar acts they or their friends engaged in); and mirroring participants' appearance and vernacular as appropriate. Validators would also observe the respondent's behavior from afar, as well as converse with peers and family. The goal was to attain insider status, and over time validators became a routine presence in study communities.

Without knowing the respondent's survey response, y^s, the validators coded an indicator of whether or not the respondent engaged in the behaviors in the two weeks prior to the survey, y^v. The authors reviewed the evidence and the coding for every case. In general, we used a relatively high standard of evidence, only coding $y^v = 1$ for a direct admission of the behavior or persuasive statements that they did not engage in the behavior.[52]

If this technique simply reproduced the errors in survey data, then the validation is little help. The key assumption is that four days of building trust and gathering extensive information, regarding just six behaviors, reduced experimenter demand and other biases correlated with treatment compared to responses during a 300-question, 90-minute questionnaire.

Nonetheless, y^v is not free from error. Appendix F reviews our approach and its limitations in more detail. Many of these limitations—the requirement of a direct admission, the disruption in people's lives, errors in recall periods, or increased social desirability bias from scrutiny—undoubtedly led to systematic errors in y^v. These errors, however, are not necessarily correlated with treatment. This is possible, for example because validators could have learned men's treatment status in conversation, and this could have biased their coding. Nonetheless, we designed the trust-building and evidentiary standards to minimize this risk.

7.2 Survey-validation differences

Of the 297 men we selected for validation, we found and validated 240 (81%).[53] Table 9 reports the means of y^s and y^v in the full sample and each treatment arm, as well as the percentage of times the two measures agree. y^s and y^v are identical about 80% of the time for sensitive measures and about 70% of the time for expenditures. As expected, however,

---

[52]The validators only witnessed or received third-party evidence of the behavior in a fifth of cases, but neither was considered sufficient evidence for a final coding. Both had to be followed by questions confirming that the respondent also engaged in the behavior in the two weeks prior to the survey. In general, we used a relatively high standard of evidence, only coding $y^v = 1$ if the validator directly observed the behavior or the respondent directly admitted it.

[53]Attrition was higher than the survey as we could not validate the behaviors of men who migrated across the country. Attrition was not correlated with treatment or baseline covariates (Blattman et al., 2015).

$\bar{y}^s < \bar{y}^v$: The average person reported 1.21 sensitive behaviors and 1.09 expenditures in validation, and 1.12 sensitive behaviors and 0.82 expenditures in the survey.

With this sample, only the underreporting of expenditures is statistically significant. We report t tests of the simple difference, $y_i^s - y_i^v$, in Appendix F. Expenditure underreporting appears to be largest in the control group, possibly because they are trying to appear more needy. Among sensitive behaviors, underreporting is generally less than 10% of the survey means, and is only statistically significant in the case of gambling. This is mainly driven by the cash only arm, who may have been reluctant to report spending the grant this way.

7.3   Is measurement error correlated with treatment?

Empirical strategy   If we believe that the validation measure is closer to the true behavior, then one way to test for bias in the survey-based treatment effects is to take the difference $y_i^s - y_i^v$, our proxy of measurement error for person i, and regress it on treatment:

$$y_i^s - y_i^v = \beta_0 + \beta_1 T + \mu_i. \tag{2}$$

If $\beta_1 < 0$ for sensitive measures, then treated men were less likely to report bad behavior, and our survey-based treatment effects may overestimate the decline in anti-social behaviors. And if $\beta_1 > 0$ for expenditures, then treated men may have over-reported their expenditures, and our survey-based treatment effects may overestimate the short-run increase in income.

With a sample of 240, we estimate we are powered to detect average under- or over-reporting of at least 14%, and error correlated with treatment of 28%.[54] Because of power concerns, we pay close attention to the sign, magnitude, and confidence interval for β_1.

Of course, the crucial assumption is that y^v is closer to the true behavior. This parallels the "no liars" and "no design effects" assumptions in list experiments. The assumption cannot be tested directly, but can only be argued on context and the quality of the approach.

We can also let misreporting vary by whether validation confirmed the behavior:

$$y_i^s = \tilde{\beta}_0 + \tilde{\beta}_1 T_i + \tilde{\beta}_2 y_i^v + \tilde{\beta}_3 \left(y_i^v \times T_i\right) + \tilde{\mu}_i. \tag{3}$$

Equation 2 is the special case where $\tilde{\beta}_2 = 1$ and $\tilde{\beta}_3 = 0$. We are interested in whether

---

[54]Our target sample of 297 was the maximum number of interviews we felt qualified validators could manage logistically. We calculated minimum detectible effects (MDEs) using a two-sided hypothesis test with 80% power at a 0.05 significance level, using baseline and block controls when calculating the R-squared statistic. We calculated an MDE for both the 0–2 expenditures index and the 0–4 sensitive behaviors index. The expenditures index had a mean of .82 in the survey and an MDE of .13 for general over- and under-reporting and .29 for a treatment effect on misreporting. The sensitive behaviors index had a mean of 1.12 in the survey and an MDE of .2 for general over- and under-reporting and .36 for any treatment effect on misreporting. We estimate that doubling the sample size would have increased power by about a third.

Table 9: Comparison of survey and qualitative validation means at endline

	Potentially sensitive behaviors					Expenditures			All
	All (0-4)	Steal	Marijuana	Gamble	Homeless	All (0-2)	Video	Phone	(0-6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. FULL SAMPLE									
Survey mean (y^s)	1.12	0.22	0.48	0.18	0.23	0.82	0.42	0.39	1.93
	(1.14)	(0.42)	(0.50)	(0.39)	(0.42)	(0.73)	(0.50)	(0.49)	(1.31)
Validation mean (y^v)	1.21	0.20	0.51	0.29	0.21	1.09	0.61	0.48	2.30
	(1.18)	(0.40)	(0.50)	(0.45)	(0.41)	(0.74)	(0.49)	(0.50)	(1.21)
% in agreement		79%	85%	72%	82%		62%	82%	
B. CONTROL GROUP									
Survey mean (y^s)	1.25	0.27	0.48	0.23	0.27	0.68	0.37	0.32	1.93
	(1.31)	(0.45)	(0.50)	(0.43)	(0.45)	(0.70)	(0.49)	(0.47)	(1.44)
Validation mean (y^v)	1.30	0.23	0.49	0.34	0.23	1.18	0.65	0.54	2.48
	(1.23)	(0.42)	(0.50)	(0.48)	(0.42)	(0.70)	(0.48)	(0.50)	(1.21)
% in agreement		80%	88%	72%	77%		47%	75%	
C. THERAPY ONLY									
Survey mean (y^s)	1.06	0.19	0.48	0.17	0.22	0.81	0.41	0.41	1.87
	(1.11)	(0.39)	(0.50)	(0.38)	(0.42)	(0.75)	(0.50)	(0.50)	(1.35)
Validation mean (y^v)	1.09	0.17	0.48	0.24	0.20	0.98	0.54	0.44	2.07
	(1.14)	(0.38)	(0.50)	(0.43)	(0.41)	(0.76)	(0.50)	(0.50)	(1.24)
% in agreement		80%	89%	74%	80%		72%	81%	
D. CASH ONLY									
Survey mean (y^s)	1.03	0.21	0.49	0.13	0.21	0.77	0.37	0.40	1.81
	(1.16)	(0.41)	(0.50)	(0.34)	(0.41)	(0.71)	(0.49)	(0.49)	(1.35)
Validation mean (y^v)	1.32	0.23	0.53	0.33	0.24	1.00	0.55	0.45	2.32
	(1.26)	(0.42)	(0.50)	(0.47)	(0.43)	(0.81)	(0.50)	(0.50)	(1.33)
% in agreement		76%	82%	74%	90%		56%	85%	
E. THERAPY + CASH									
Survey mean (y^s)	1.13	0.22	0.48	0.21	0.22	0.98	0.54	0.44	2.11
	(0.98)	(0.42)	(0.50)	(0.41)	(0.42)	(0.73)	(0.50)	(0.50)	(1.11)
Validation mean (y^v)	1.11	0.19	0.52	0.24	0.16	1.17	0.70	0.48	2.29
	(1.11)	(0.40)	(0.50)	(0.43)	(0.37)	(0.68)	(0.46)	(0.50)	(1.05)
% in agreement		81%	83%	68%	81%		71%	87%	
Observations	239	238	238	238	239	239	238	239	239

Notes: The table reports the means (standard deviations) of the survey and the qualitatively validated measures for the full sample and by treatment arm. "% in agreement" is the percentage of respondents for whom the survey indicator equals the qualitatively validated indicator.

$\tilde{\beta}_1 \neq 0$ and $\tilde{\beta}_1 + \tilde{\beta}_3 \neq 0$. The disadvantage of this more flexible form is statistical power, especially with three treatment arms.[55] We are also interested in correcting for the average bias in survey-based treatment effects, which we get from β_1 from equation 2. But the more flexible form provides insight into the patterns of measurement error. For instance, if there is a general desirability bias in the survey, then $\tilde{\beta}_0 + \tilde{\beta}_2 < 1$, whereas if underreporting is concentrated among men who commit crimes and were treated, then $\tilde{\beta}_1 + \tilde{\beta}_3 < 0$.[56]

Results for sensitive behaviors We estimate equations 2 and 3 in Table 10. For sensitive behaviors, almost none of the coefficients on treatment indicators or interactions are statistically significant. We see little evidence of the therapy inducing a desirability bias, and indeed the effects run in the opposite direction.

Indeed, looking at the index of four sensitive measures (Panel (a), Column 5), β_1 is actually greater than zero for therapy and therapy plus cash, implying that the impacts of therapy are, if anything, larger than the survey data imply. Appendix F displays these updated treatment effects. For example, using survey data alone, the treatment effect (standard error) of therapy and cash on the sensitive behaviors index is -0.4 (0.09), a 36% decrease. The results from Panel (a), Column 5 suggest that the adjusted treatment effect should be -0.516 (.194), significant at the 1% level.

The results of the more flexible regression in Panel (b), Column 5 shows that these averages conceal important heterogeneity. Treated men who we think did not engage in the sensitive behaviors tend to over-report them ($\tilde{\beta}_1^{Both} > 0$), and treated men engaged in the sensitive behaviors seem to under-report them ($\tilde{\beta}_1^{Both} + \tilde{\beta}_3^{Both} < 0$).[57]

Results for expenditures All treatment arms associated with a roughly 0.3 increase in our proxy for measurement error (Panel (a), Column 8). There is underreporting across all arms, but it is greatest in the control group. This could have implications for one of our main findings, on income. Using survey data, the treatment effect of cash only on the 2-item expenditure index is 0.08 (0.052), which is consistent with the short run increase in consumption we observed among cash recipients. But adjusting for observed measurement error, the adjusted treatment effect is -0.205 (0.143).

---

[55]With 240 observations in total, each parameter is estimated off of roughly 30 observations, putting us on a steep part of the power curve.

[56]Moreover, if men honestly report crime in the survey then $\tilde{\beta}_0$ should be close to zero and $\tilde{\beta}_2$ should be close to 1. Appendix F derives and interprets these regressions in more detail.

[57]Also, note that, on average, $\tilde{\beta}_0 > 0$, $\tilde{\beta}_2 < 1$, and $\tilde{\beta}_0 + \tilde{\beta}_2 < 1$ for sensitive measures (Column 5). This is consistent with what we observe in Table 9: slight survey underreporting of sensitive behaviors, and 20–30% non-correspondence between survey and validated measures.

Table 10: Estimates of the correlation between treatment and measurement error

(a) Constrained, with block fixed effects (Equation 2)

	Dependent variable (N=239)							
Covariate	$y^s - y^v$, Sensitive behaviors					$y^s - y^v$, Expenditures		
	Stealing	Marijuana	Gambling	Homeless	All (0–4)	Video Club	Phone Charging	All (0–2)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_o (Constant)	0.087	0.555	-0.638	-0.096	-0.084	-1.330	-0.667	-1.987
	[.124]	[.395]	[.363]*	[.098]	[.803]	[.137]***	[.321]**	[.384]***
β_1								
Therapy	-0.005	-0.012	0.122	0.056	0.158	0.276	0.240	0.504
	[.092]	[.064]	[.111]	[.105]	[.209]	[.112]**	[.089]***	[.145]***
Cash	0.045	-0.025	-0.076	0.042	-0.002	0.051	0.142	0.182
	[.102]	[.067]	[.093]	[.079]	[.204]	[.116]	[.083]*	[.138]
Both	0.058	-0.011	0.185	0.159	0.383	0.222	0.139	0.346
	[.094]	[.081]	[.107]*	[.100]	[.228]*	[.116]*	[.089]	[.127]***

(b) Unconstrained, with block fixed effects (Equation 3)

	Dependent variable (N=239)							
Covariate	y^s, Sensitive behaviors					y^s, Expenditures		
	Stealing	Marijuana	Gambling	Homeless	All (0–4)	Video Club	Phone Charging	All (0-2)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\tilde{\beta}_0$ (Constant)	0.073	0.613	-0.179	-0.096	0.157	-0.472	-0.162	-0.788
	[.088]	[.381]	[.091]*	[.087]	[.599]	[.160]***	[.290]	[.396]**
$\tilde{\beta}_1$								
Therapy	-0.050	-0.092	0.066	0.067	0.215	-0.061	0.112	0.113
	[.076]	[.082]	[.083]	[.093]	[.235]	[.131]	[.076]	[.207]
Cash	0.008	0.059	-0.068	-0.012	0.123	-0.044	0.047	0.163
	[.072]	[.095]	[.072]	[.069]	[.219]	[.148]	[.083]	[.225]
Both	-0.009	-0.044	0.083	0.095	0.455	0.032	0.032	-0.121
	[.077]	[.104]	[.078]	[.089]	[.274]*	[.147]	[.080]	[.216]
$\tilde{\beta}_2(y^v)$	0.376	0.667	0.249	0.382	0.583	0.097	0.488	0.333
	[.152]**	[.110]***	[.107]**	[.155]**	[.113]***	[.125]	[.098]***	[.131]**
$\tilde{\beta}_3$								
Therapy$\times y^v$	-0.091	0.115	-0.097	-0.094	-0.169	0.462	0.194	0.318
	[.221]	[.134]	[.161]	[.221]	[.135]	[.170]***	[.148]	[.173]*
Cash$\times y^v$	-0.155	-0.108	0.026	0.211	-0.115	0.097	0.166	-0.031
	[.204]	[.144]	[.159]	[.198]	[.152]	[.181]	[.156]	[.177]
Both$\times y^v$	-0.098	-0.029	-0.246	-0.012	-0.318	0.385	0.232	0.473
	[.203]	[.147]	[.177]	[.234]	[.155]**	[.180]**	[.142]	[.174]***

Notes: The table reports the degree and direction of bias in our treatment effects. In Figure A, we assume that our measurement error does not vary by whether or not the individual engages in the behavior, which allows for a simple way to use β_2 to adjust our ITT estimates. In Figure B, we relax this assumption and let the measurement error vary by behavior and treatment arm at the cost of reduced statistical power.

Interpretation   Our qualitative work suggests two explanations. The men have been members of a subculture where drugs, crime, and gambling are commonplace, and admitting to the behaviors in a survey carries little stigma. Speculatively, therapy may have accustomed men to talking about these behaviors or reduced stigma. As for expenditures, control men may have acted strategically, trying to appear poorer in the hopes they would be eligible for assistance. We discuss implications for our conclusions in the following section.

8   Discussion

8.1   Lessons from the cash transfer

First, these supposedly undisciplined men largely invested and saved a grant. Even accounting for the underreporting we see in gambling and other expenditures, little of the grants seem to have been spent on temptation goods.[58] In the short run, men used the cash for petty trade, earning returns to capital of at least 26%.[59] Unfortunately we cannot say whether the cash grant passed a cost-benefit test in private monetary returns alone. Caution is also warranted, because of the evidence that the control group underreported expenditures.

Second, crime fell as business income rose. The income gain had little effect on aggression, but those who received the cash reduced stealing by a third. This is consistent with rural ex-combatants in Liberia, who shifted away from illicit activities when a program raised their farm productivity (Blattman and Annan, 2015).

Third, these investments and income gains disappeared within a year, in part due to poor property rights protections.[60] The men's homes and neighborhoods were highly insecure. Extrapolating from reports of burglary and theft at each endline (from Table 4), men in our sample experienced a theft or robbery roughly eight times in the year after the grant. While treated men were no more likely to experience theft, they had more to lose, especially their savings and investment in nascent businesses.

Nonetheless, the fact that cash was well-used is important, since concerns about temptation spending restrain political support for cash-based welfare programs. The men received a few months worth of income, and basic consumption—especially basic shelter and food—improved for about that length of time. This is important.

---

[58]Evans and Popova (2015) see the same result in 19 other cash transfer programs, but it's striking to see the same with this extreme group.

[59]For instance, the impact on earnings ($8.25 a month) represents a monthly return of 4.1% on the $200 grant, while the impact on non-durable consumption ($48 per month) represents a monthly return of 24%. While there are reasons these figures might overstate returns, recall that men only invested about 60% in the month after the grant, implying returns on actual investment are probably higher.

[60]This contrasts with a literature showing that poor young people in Africa invest grants and increase incomes (Fafchamps et al., 2014; Blattman et al., 2014, 2015).

Future research should study how to sustain the economic effects of cash. It may be that helping people relocate to better quality neighborhoods or enhance personal security, or providing the information and means to gain necessary licenses or protection from security forces would reduce expropriation. Alternately, programs can try to provide crude insurance. It is possible that regular cash transfers would stimulate enterprise development more than the one-time transfer we study (Bianchi and Bobba, 2013; Karlan et al., 2015).

8.2   Lessons from behavior change

First, self control skills seem to be malleable in adults and respond to investments such as group therapy, at least in this subject group and context. The most significant signs of change were in impulsivity and reward motivation. While psychologists have tended to consider these as stable traits, most therapies for extreme risky behavior try to teach tactics for shifting impulsive behavior. Indeed, our results echo the effects of adolescent CBT programs in Chicago that target similar automatic behaviors (Heller et al., 2015).[61] Conscientiousness and grit receive more emphasis in the economics literature on noncognitive skills, but these seem to be less responsive to the therapy, and are possibly less malleable.

Second, there appears to be something to the least standard aspect of the therapy: the focus on changing social identity, and with it the values and norms to which the men subscribe. Qualitatively, the changes in appearance, in community regard, and the exposure to new places and situations seem to have been particularly important. So was the identity of the NEPI facilitators, and the fact that they modeled this image change. This change has a basis in the theory underlying CBT: positive interactions challenged respondents' negative beliefs about themselves, and reinforced their self image as more responsible, mainstream members of society. There are reasons for caution, however. We did not measure self image directly, but rather only have evidence of self-reported value change. It will be important for future rounds or future trials to improve measurement.

Third, we did not see large, statistically significant, long term effects of therapy on various secondary mechanisms, including the quality of social networks. These were part

---

[61]The CBT approach may also have been important, as a randomized trial of another NEPI intervention that did not follow these principles had no effect on attitudes, values, or behaviors, despite having some of the same facilitators and trainers (Blattman and Annan, 2015) Prior to this study, NEPI was hired by an international non-profit to conduct a residential group therapy program for rural ex-combatants, in tandem with agricultural training. While there was overlap in curriculum with STYL, the residential therapy had a more diverse array of topics (including dealing with trauma and civic education); did not formally include homework or follow-up or exposure to new social interactions; and socialized young men in an artificial environment outside their home. The subjects were considered high-risk but had lower rates of crime, drugs, and violence than their urban STYL counterparts. Given differences in design, facilitators, and subjects, we cannot causally attribute the absence of impacts on antisocial behavior to the therapeutic approach, but the difference is consistent with the theory underlying CBT.

of the STYL curriculum and aims, although not central ones. There are several reasons we may not see a change, including: poor measurement; lower malleability of these traits; or the specific design and content of the therapy. In psychology, efficacy trials such as this one are typically followed by further trials that try to identify the "active ingredients", by varying modules and methods. This seems like a fruitful area for research.

Understanding the cash–therapy interaction

We did not expect that the effects of therapy would persist only when cash was received as well. Our theory predicted that the two interventions should have a larger effect only if cash raises earnings permanently, which was not the case.

Our qualitative evidence and psychological theory, however, suggest a hypothesis for testing in future trials: that receiving cash was akin to an extension of therapy, in that it provided more time for the men to independently practice and reinforce their changed skills, image, and behaviors. The therapy was brief—just eight weeks long. It helped men change their intentions, image and behavior, and provided almost daily commitment and reinforcement. After eight weeks the men who received therapy alone had to contend with their usual economic and peer pressures. The grant, however, provided some men with the cash they needed to maintain their new image—to avoid homelessness, to feed themselves, and to continue to dress well. They had no immediate financial need to return to crime.

The men could also do something consistent with their new image and skills: execute plans for a business. This was a source of practice and reinforcement of their newfound skills and identity. It was also a form of performance, to themselves as well as their family and neighbors, who could see the men engage in legitimate business. Our qualitative interviews also suggested that the cash helped men to survive shocks. In this way, the grant may have parallels to "booster sessions" commonly used in therapy. A small body of experimental research on CBT for aggression or substance abuse indicates that follow-up therapy sessions weeks or months after the intervention improves long term outcomes (e.g. Lochman, 1992).

Caution is warranted. We cannot reject the hypothesis, for instance, that positive reinforcement from winning a grant was enough to reinforce therapy. In future, a comparison of extended therapy to shorter therapy plus cash would offer a more direct test.

Nonetheless, high short-run returns to capital and sustained social spillovers suggests that the combination of cash and therapy had promising returns. Since the private returns to the grant were temporary, however, the cost effectiveness rides mostly on the social benefits from roughly one fewer crime per week per person. These social returns are unknown. If these social returns are greater than $20 or $25 per crime, however, the STYL program is a promising investment on basis of crime reduction alone.

8.3 Generalizability

For several reasons this approach has promise beyond Liberia. First, the therapy was adapted from U.S.-based CBT programs, suggesting that adaptability to other contexts is feasible. Second, we kept the intervention low-cost and created a publicly-available manual, curriculum, and training guidelines to ease adaptation and replication. Third, with time it should be possible to develop qualified and effective facilitators in other countries, not least because there are established methods for training counselors in CBT; general levels of education (and the number of social workers) are greater in most other countries; and new facilitators should emerge among graduates of the program, as with STYL.

The theory and results are also strikingly consistent with comparable U.S. programs and best practice. The attention to noncognitive skill change and self image, the targeting of the highest-risk men, as well as the non-residential nature of the therapy, correspond closely to best practice in criminal rehabilitation in U.S. correctional institutions (Andrews et al., 1990; Lipsey, 2009). The 40–50% falls in antisocial behaviors we observe are similar in proportion to the falls in arrests documented in Tennessee and Chicago (Little et al., 1994; Heller et al., 2015). Moreover, as in Chicago, the effects of therapy alone were temporary.

Other U.S. work suggests that employment can be complementary to social and emotional counseling (Heller, 2014). In low-income countries, however, where most employment programs will involve self-employment, property security and risk are important scope considerations. Cash transfers in other poor countries have generally led to higher and more persistent incomes, in part because the gains are not stolen. So the STYL program could arguably work even better in places with more secure property rights.

There are limits to generalizability of course. For instance, there were no gangs or armed groups vying for men in our sample. CBT-based approaches may be most effective against disorganized, impulsive crime and violence rather than organized crime. There is also selection onto the street, and a country which has experienced many negative shocks (such as Liberia) might have more high-potential young men who need only a little help to regress to the mean. On the other hand, our evidence from dropouts suggests that the most antisocial men stay, and the program is most effective with them. These limits are speculative without further testing, however, and replication and experimentation seem more than warranted given the results of these efficacy trials in Liberia, Chicago, and elsewhere.

References

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *Quarterly Journal of Economics 115*(3), 715–753.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Personality Psychology and Economics. In *Handbook of the economics of education*, Volume 4(1), pp. 1–181. Elsevier.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Andrews, D. A., I. Zinger, R. D. Hoge, J. Bonta, P. Gendreau, and F. T. Cullen (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology 28*(3), 369–404.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond.* New york: Guilford Press.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012). The Relationship Between Economic Preferences and Psychological Personality Measures. *Annual Review of Economics 4*, 453–78.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy 76*, 169–217.

Bénabou, R. and J. Tirole (2004). Willpower and personal rules. *Journal of Political Economy 112*(4), 848–886.

Bernard, T., S. Dercon, K. Orkin, and A. S. Taffesse (2014). The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia. *Working paper*.

Bianchi, M. and M. Bobba (2013). Liquidity, Risk, and Occupational Choices. *Review of Economic Studies 80*(2), 491–511.

Blattman, C. and J. Annan (2015). Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State. *forthcoming in American Political Science Review*.

Blattman, C., J. Annan, E. P. Green, C. Lehmann, and J. Jamison (2015). The returns to microenterprise development among the ultra-poor: A field experiment in post-war Uganda. *forthcoming in American Economic Journal: Applied Economics*.

Blattman, C., N. Fiala, and S. Martinez (2014). Generating skilled employment in developing countries: Experimental evidence from Uganda. *Quarterly Journal of Economics 129*(2), 697–752.

Blattman, C., J. Jamison, T. Koroknay-Palicz, K. Rodrigues, and M. Sheridan (2015). Measuring the measurement error: A method to qualitatively validate survey data. *forthcoming in Journal of Development Economics*.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. t. Weel (2008, September). The Economics and Psychology of Personality Traits. *Journal of Human Resources 43*(4), 972–1059.

Christensen, M. M. and M. Utas (2008). Mercenaries of Democracy: the 'Politricks' of Remobilized Combatants in the 2007 General Elections, Sierra Leone. *African Affairs 107*(429), 1–25.

Coopersmith, S. (1967). *The antecedents of self-esteem.* San Francisco: Consulting Psychologists Press.

Costa, P. T. and R. R. McCrae (1997). Stability and change in personality assessment: the revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment 68*(1), 86–94.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Del Vecchio, T. and K. D. O'Leary (2004). Effectiveness of anger treatments for specific anger problems: a meta-analytic review. *Clinical psychology review 24*(1), 15–34.

Draca, M. and S. Machin (2015). Crime and Economic Incentives. *Annual Review of Economics 7*, 389–408.

Duckworth, A. and P. Quinn (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment 91*, 166–174.

Duckworth, A. L. and R. Schulze (2011). A Meta-Analysis of Convergent Validity Evidence for Self-Control Measures. *Journal of Personality Assessment 45*(3), 259–268.

Evans, D. K. and A. Popova (2015). Cash Transfers and Temptation Goods. *forthcoming in Economic Development and Cultural Change*.

Fafchamps, M., D. J. McKenzie, S. Quinn, and C. Woodruff (2014). When is capital enough to get female microenterprises growing? Evidence from a randomized experiment in Ghana. *Journal of Development Economics 106*(1), 211–226.

Ghosal, S., S. Jana, A. Mani, and S. Mitra (2015). Sex Workers, Self-Image and Stigma: Evidence from Kolkata Brothels. *Working paper*.

Gottfredson, M. R. and T. Hirschi (1990). *A general theory of crime.* Palo Alto: Stanford University Press.

Haushofer, J. and J. Shapiro (2013). Welfare Effects of Unconditional Cash Transfers: Evidence from a Randomized Controlled Trial in Kenya. *Working paper*.

Heckman, J. J. and T. Kautz (2014). Fostering and measuring skills: Interventions that improve character and cognition. *The Myth of Achievement Tests: The FED and the Role of Character in the American Life*, 293–317.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Heller, S. B. (2014, December). Summer jobs reduce violence among disadvantaged youth. *Science 346*(6214), 1219–1223.

Heller, S. B., H. A. Pollack, R. Ander, and J. Ludwig (2013). Preventing Youth Violence and Dropout: A Randomized Field Experiment. *NBER Working Paper 19104*.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, and H. A. Pollack (2015). Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago. *NBER Working Paper No. 21178*.

Hill, P. L., B. W. Roberts, J. T. Grogger, J. Guryan, and K. Sixkiller (2011). Decreasing Delinquency, Criminal Behavior, and Recidivism by Intervening on Psychological Factors Other than Cognitive Ability: A Review of the Intervention Literature. *NBER Working Paper 16698*.

Jolls, C., C. R. Sunstein, and R. Thaler (1998). A behavioral approach to law and economics. *Stanford Law Review 50*(5), 1471–1550.

Karlan, D., R. Knight, and C. Udry (2015). Consulting and capital experiments with microenterprise tailors in ghana. *Journal of Economic Behavior and Organization 118*, 281–302.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Langberg, J. M., S. P. Becker, J. N. Epstein, A. J. Vaughn, and E. Girio-Herrera (2013). Predictors of Response and Mechanisms of Change in an Organizational Skills Intervention for Students with ADHD. *Journal of child and family studies 22*(6), 1000–1012.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and offenders 4*(2), 124–147.

Little, G., K. Robinson, and K. Burnette (1994). Treating offenders with cognitivebehavioral therapy: 5-year recidivism outcome data on MRT. *Cognitive-Behavioral Treatment Review 3*, 1–3.

Lochman, J. E. (1992). Cognitive-behavioral intervention with aggressive boys: three-year follow-up and preventive effects. *Journal of consulting and clinical psychology 60*(3), 426–432.

Maruna, S. and K. Roy (2007, February). Amputation or Reconstruction? Notes on the Concept of "Knifing Off" and Desistance From Crime. *Journal of Contemporary Criminal Justice 23*(1), 104–124.

Millenky, M., D. Bloom, S. Muller-Ravett, and J. Broadus (2012). Staying on course: Three-year results of the National Guard Youth ChalleNGe evaluation. Technical report, MDRC, Washington DC.

Mischel, W., Y. Shoda, and M. I. Rodriguez (1989). Delay of gratification in children. *Science 244*(4907), 933–938.

Nagin, D. S. and G. Pogarsky (2004). Time and punishment: Delayed consequences and criminal behavior. *Journal of Quantitative Criminology 20*(4), 295–317.

Nelson, C. A., C. H. Zeanah, N. A. Fox, P. J. Marshall, A. T. Smyke, and D. Guthrie (2007). Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project. *Science 318*(5858), 1937–1940.

Pearson, F. S., D. S. Lipton, C. M. Cleland, and D. S. Yee (2002). The Effects of Behavioral/Cognitive-Behavioral Programs on Recidivism. *Crime & Delinquency 48*(3), 476–496.

Ponniah, K. and S. D. Hollon (2008). Empirically supported psychological interventions for social phobia in adults: a qualitative review of randomized controlled trials. *Psychological medicine 38*(1), 3–14.

Raine, A., K. A. Dodge, R. Loeber, L. Gatzke-Kopp, D. Lynam, C. Reynolds, M. Stouthamer-Loeber, and J. Liu (2006). The Reactive-Proactive Aggression Questionnaire: Differential Correlates of Reactive and Proactive Aggression in Adolescent Boys. *Aggressive Behavior 32*, 159–171.

Republic of Liberia (2012). Liberia: Poverty Reduction Strategy. Technical report, Government of Liberia, Monrovia.

Robinson, T. E. and K. C. Berridge (2000). The psychology and neurobiology of addiction: an incentive-sensitization view. *Addiction 95 (2)*, S91–117.

Saini, M. (2009). A meta-analysis of the psychological treatment of anger: developing guidelines for evidence-based practice. *The journal of the American Academy of Psychiatry and the Law 37*(4), 473–488.

Sapp, S. G. and W. J. Harrod (1993). Reliability and validity of a brief version of Levenson's locus of control scale. *Psychological Reports 72*, 539–550.

Schochet, P. Z., J. Burghardt, and S. McConnell (2008). Does Job Corps work? Impact findings from the National Job Corps Study. *American Economic Review 98*(5), 1864–1886.

Spinella, M. (2007). Normative data and a short form of the Barratt Impulsiveness Scale. *The International journal of neuroscience 117*(3), 359–368.

Vigil, J. D. (2003). Urban violence and street gangs. *Annual Review of Anthropology 32*, 225–242.

Wilson, D. B., L. A. Bouffard, and D. L. Mackenzie (2005). A Quantitative Review of Structured, Group-Oriented, Cognitive-Behavioral Programs for Offenders. *Criminal Justice and Behavior 32*(2), 172–204.

Wood, E. J. (2008). The social processes of civil war: The wartime transformation of social networks. *Annual Review of Political Science 11*, 539–561.

Appendix for online publication

A   Baseline sample

A.1   Full summary statistics and balance tests

Table A.1 expands the balance table in the main paper for the full set of baseline covariates available and used in the treatment effects regressions.[1] Column 1 reports the sample mean for each covariate, and Columns 2 to 7 report the coefficients and p values on treatment indicators from ordinary least squares (OLS) regressions of each baseline covariate on three treatment indicators (one for assignment to each treatment arm) controlling for block fixed effects. Column 8 reports the p value from a joint test of significance of the three coefficients. Finally, at the base of the table we report the p value from a test of joint significance of all covariates from an OLS regression of each treatment indicator on all covariates (including that treatment group and the control group alone).

Of 171 coefficients (57 covariates and 3 treatment arms), 10 (5.8%) have a $p < .05$, and 26 (15.2%) have a $p < .1$. Within treatment arms the covariates are not jointly significant, as seen from the joint test reported at the base of the table. Furthermore, 9 (15.8%) of the tests of joint significance have a $p < .1$.

Table A.2 repeats the same balance analysis for the 947 subjects interviewed at endline. Of 171 coefficients (57 covariates and 3 treatment arms), 13 (7.6%) have a $p < .05$, and 20 (11.6%) have a $p < .1$. Within treatment arms the covariates are not jointly significant, as seen from the joint test reported at the base of the table. Furthermore, 12 (21.1%) of the tests of joint significance have a $p < .1$.

Overall, therefore, there is minor imbalance. We control for all baseline covariates in all treatment effects regressions in the paper to account for this.

A.2   Neighborhoods and recruitment

Table A.3 describes each of the study neighborhoods where we recruited, along with population estimates. We report the estimates of the number of all adult males, as well as our low-end estimates of the number of target males in each neighborhoods—men 18 to 35 in the bottom decile of income.

A.3   Tracking and attrition

We achieved tracking rates of roughly 93% over a year.[2] Given that this was such a transient population, we took special measures to minimize attrition.

---

[1] We maintained the Phase 1 baseline survey for all Phases for the sake of consistency and completeness.

[2] Rates of 80, 90 or even 95 percent are not uncommon in developing country field experiments and panel surveys. For example, the Indonesia Family Life Survey reached 94% of households and 91% of target individuals after four years. The Kenyan Life Panel Survey made contact with 84 percent of target respondents over a seven-year period. Similarly, in the US, researchers were able to reach 98% of the Perry Pre-school children at age 19 and 95% at age 27. One reason is that a small sample is easier to track intensively. Another reason is that enumerator wages are lower in Liberia in the U.S. and this means that intensive sleuthing and tracking is affordable.

Table A.1: Baseline statistics and balance test

Baseline covariate	Sample Mean (1)	Assigned therapy Coeff. (2)	Assigned therapy p value (3)	Assigned cash Coeff. (4)	Assigned cash p value (5)	Assigned both Coeff. (6)	Assigned both p value (7)	F-Test p value (8)
		Test of randomization balance (continued) (N=999)						
Age	25.40	-0.16	0.68	0.19	0.59	-0.18	0.68	0.18
Married or partnered	0.16	-0.03	0.65	-0.04	0.67	0.04	0.76	0.93
# of partners	0.53	0.06	0.43	0.12	0.17	-0.21	0.12	0.11
# of children<15 in household	2.20	-0.59	0.07	-0.50	0.19	0.62	0.30	0.33
Sees family often	2.36	0.13	0.00	0.25	0.02	-0.30	0.01	0.01
Muslim	0.10	0.02	0.64	0.00	0.96	0.01	0.87	0.18
Years of schooling	7.72	-0.19	0.68	0.04	0.95	-0.01	0.99	0.55
Currently in school	0.06	-0.03	0.08	-0.03	0.13	0.04	0.11	0.16
Literacy index (0-2)	1.21	0.14	0.11	0.13	0.30	-0.27	0.08	0.12
Math score (0-5)	2.79	-0.10	0.25	-0.03	0.85	-0.15	0.39	0.89
Health index (0-6)	4.87	-0.09	0.11	-0.19	0.17	0.31	0.15	0.28
Has any disabilities	0.08	0.04	0.29	0.00	1.00	-0.04	0.48	0.19
Depression index (0-17)	7.09	0.18	0.41	-0.01	0.97	-0.11	0.80	0.45
Distress index (0-21)	7.51	0.16	0.43	0.00	0.99	-0.40	0.31	0.40
Relations to commanders index (0-4)	0.45	0.00	0.93	0.07	0.42	-0.06	0.55	0.72
Ex-combatant	0.41	0.06	0.09	0.08	0.11	-0.09	0.12	0.15
War experiences index (0-12)	5.85	0.38	0.24	0.19	0.48	-0.73	0.12	0.32
Weekly cash earnings (USD)	17.02	-1.89	0.03	-4.85	0.03	5.48	0.00	0.02
Summary index of income, z-score	0.00	-0.22	0.05	-0.12	0.48	0.26	0.21	0.07
Homeless in past two weeks	0.25	-0.01	0.82	0.00	0.93	-0.02	0.74	0.33
# of days slept hungry, last 7 days	1.26	0.25	0.10	0.28	0.05	-0.32	0.09	0.14
Savings stock (USD)	33.83	-10.08	0.26	-12.74	0.31	15.71	0.31	0.53
Can get loan of 50 USD	0.52	-0.02	0.62	-0.05	0.32	0.04	0.51	0.57
Can get loan of 300 USD	0.11	-0.03	0.27	-0.03	0.34	0.06	0.07	0.13
Hours in illicit activities	13.55	1.21	0.68	-0.86	0.67	0.06	0.99	0.14
Hours/week in agriculture	0.36	0.34	0.26	-0.10	0.35	0.13	0.84	0.01
Hours/week in low-skill wage labor	19.39	0.54	0.88	1.24	0.73	-0.43	0.90	0.94
Hours/week in low-skill business	11.53	0.16	0.92	-1.53	0.60	5.76	0.13	0.50
Hours/week in high-skill work	1.51	-0.05	0.91	0.94	0.03	0.11	0.85	0.01
Years of experience in agriculture	0.78	-0.21	0.29	-0.34	0.07	0.25	0.32	0.15
Years experience in non-agricultural business	2.96	-0.35	0.36	-0.80	0.05	0.97	0.08	0.04
Years experience in high-skill work	0.96	-0.29	0.13	-0.27	0.41	0.62	0.12	0.02
Sells drugs	0.20	0.01	0.69	0.00	0.92	0.00	0.93	0.92
Drinks alcohol	0.75	0.08	0.19	0.07	0.23	-0.07	0.23	0.31
Uses marijuana	0.59	0.12	0.02	0.09	0.02	-0.14	0.01	0.01
Uses marijuana daily	0.44	0.08	0.14	0.04	0.12	-0.09	0.21	0.34
Use hard drugs	0.26	-0.01	0.82	0.02	0.59	-0.01	0.82	0.83
Uses hard drugs daily	0.15	-0.04	0.21	0.02	0.52	0.01	0.90	0.37

Continued on following page.

Table A.1 (continued): Baseline statistics and balance test

	Sample	Test of randomization balance (N=999)						
		Assigned therapy		Assigned cash		Assigned both		F-Test
Baseline covariate	Mean	Coeff.	p value	Coeff.	p value	Coeff.	p value	p value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Committed theft/robbery in past 2 weeks	0.53	0.05	0.51	0.01	0.64	-0.02	0.66	0.80
Number of nonviolent stealing incidents	5.09	-0.36	0.58	-0.46	0.68	0.39	0.70	0.88
Number of felony stealing incidents	0.43	0.06	0.77	0.17	0.60	-0.17	0.67	0.86
Disputes and fights in past 2 weeks (0-9)	2.16	0.14	0.80	0.33	0.63	-0.68	0.25	0.68
Aggressive behaviors (mean of 19), z-score	0.00	0.05	0.66	0.13	0.22	-0.23	0.09	0.23
Conscientiousness index (0-24)	15.36	-0.05	0.74	-0.22	0.32	-0.01	0.98	0.05
Neuroticism index (0-21)	12.09	-0.07	0.77	0.18	0.64	0.11	0.85	0.34
Grit index (0-21)	13.76	0.07	0.59	-0.08	0.83	0.00	0.99	0.24
Reward responsiveness index (0-24)	14.75	-0.15	0.48	-0.03	0.95	-0.25	0.70	0.95
Locus of control index (0-24)	14.45	-0.09	0.77	-0.43	0.15	0.45	0.29	0.00
Impulsiveness index (0-21)	9.39	0.39	0.38	0.18	0.66	-0.88	0.10	0.33
Self esteem index (0-24)	13.47	-0.08	0.78	-0.11	0.65	0.12	0.75	0.89
Patience in game play index (0-6)	4.11	0.08	0.48	-0.07	0.77	0.03	0.94	0.83
Time inconsistency in game play index (0-6)	3.27	-0.22	0.03	-0.05	0.62	0.13	0.34	0.01
Risk aversion index (0–3)	1.56	-0.02	0.89	-0.05	0.56	0.09	0.43	0.61
Self-reported patience (mean of 7), z-score	0.00	-0.08	0.62	-0.13	0.25	0.15	0.42	0.33
Declared Risk Appetite (mean of 6), z-score	0.00	0.01	0.94	-0.02	0.88	-0.10	0.65	0.94
Cognitive ability (z-score)	0.00	0.15	0.05	0.14	0.20	-0.29	0.01	0.04
Executive function (z-score)	0.00	0.07	0.18	0.10	0.45	-0.25	0.06	0.16
R-Squared		0.17		0.11		0.33		
p Value on F-statistics on all covariates		0.53		0.66		0.25		

Notes: Column (1) reports the sample mean. A small number of missing values are imputed at the median. Columns (2)-(7) report the coefficients and p values from ordinary least squares regressions of each baseline covariate on three indicators, one for assignment to each treatment arm, controlling for block fixed effects. Column (8) reports the p value from a joint test of statistical significance of all three treatment indicators.

Table A.2: Baseline statistics and balance test for endline respondents

| Baseline covariate | Sample Mean (1) | Test of randomization balance (continued) (N=947) | | | | | | F-Test |
| | | Assigned therapy | | Assigned cash | | Assigned both | | |
		Coeff. (2)	p value (3)	Coeff. (4)	p value (5)	Coeff. (6)	p value (7)	p value (8)
Age	25.35	-0.23	0.49	0.05	0.89	-0.02	0.96	0.17
Married or partnered	0.16	-0.02	0.75	-0.03	0.70	0.03	0.83	0.90
# of partners	0.53	0.06	0.43	0.12	0.20	-0.20	0.14	0.20
# of children<15 in household	2.23	-0.58	0.06	-0.47	0.19	0.64	0.27	0.28
Sees family often	2.36	0.10	0.06	0.22	0.01	-0.27	0.02	0.02
Muslim	0.09	0.02	0.56	0.01	0.60	0.00	0.95	0.75
Years of schooling	7.73	-0.30	0.51	-0.06	0.89	0.14	0.79	0.54
Currently in school	0.06	-0.03	0.09	-0.03	0.13	0.04	0.16	0.20
Literacy index (0-2)	1.21	0.08	0.33	0.08	0.47	-0.20	0.14	0.30
Math score (0-5)	2.79	-0.18	0.15	-0.09	0.66	-0.08	0.71	0.76
Health index (0-6)	4.85	-0.07	0.30	-0.19	0.21	0.28	0.25	0.43
Has any disabilities	0.07	0.04	0.29	0.00	0.97	-0.05	0.49	0.22
Depression index (0-17)	7.08	0.03	0.90	-0.07	0.82	0.06	0.90	0.88
Distress index (0-21)	7.49	-0.01	0.98	-0.10	0.80	-0.15	0.74	0.82
Relations to commanders index (0-4)	0.45	0.00	0.99	0.04	0.61	-0.04	0.74	0.88
Ex-combatant	0.41	0.06	0.11	0.07	0.19	-0.08	0.19	0.25
War experiences index (0-12)	5.86	0.34	0.37	0.01	0.96	-0.61	0.22	0.43
Weekly cash earnings (USD)	16.90	-1.94	0.11	-4.33	0.03	4.70	0.00	0.01
Summary index of income, z-score	-0.01	-0.20	0.03	-0.11	0.41	0.25	0.14	0.06
Homeless in past two weeks	0.25	0.00	0.90	0.01	0.77	-0.03	0.62	0.26
# of days slept hungry, last 7 days	1.27	0.26	0.07	0.30	0.02	-0.37	0.04	0.07
Savings stock (USD)	32.54	-7.32	0.34	-7.11	0.49	10.34	0.42	0.60
Can get loan of 50 USD	0.51	-0.02	0.71	-0.04	0.37	0.04	0.49	0.58
Can get loan of 300 USD	0.10	-0.02	0.42	-0.02	0.50	0.05	0.28	0.45
Hours in illicit activities	13.22	0.61	0.80	-0.69	0.68	0.03	0.99	0.24
Hours/week in agriculture	0.37	0.35	0.33	-0.16	0.21	0.14	0.86	0.02
Hours/week in low-skill wage labor	19.34	0.46	0.91	0.78	0.84	-0.28	0.95	0.99
Hours/week in low-skill business	11.73	0.18	0.90	-1.56	0.60	5.54	0.15	0.53
Hours/week in high-skill work	1.46	-0.19	0.76	1.12	0.03	0.46	0.57	0.02
Years of experience in agriculture	0.74	-0.20	0.28	-0.30	0.12	0.27	0.25	0.31
Years experience in non-agricultural business	3.03	-0.41	0.24	-0.86	0.03	0.98	0.07	0.02
Years experience in high-skill work	0.93	-0.15	0.13	-0.04	0.87	0.44	0.11	0.03
Sells drugs	0.20	0.00	0.97	0.00	0.99	0.02	0.69	0.96
Drinks alcohol	0.76	0.06	0.28	0.07	0.30	-0.07	0.18	0.35
Uses marijuana	0.59	0.11	0.03	0.08	0.06	-0.14	0.02	0.04
Uses marijuana daily	0.44	0.07	0.20	0.04	0.18	-0.08	0.25	0.37
Use hard drugs	0.26	-0.02	0.57	0.01	0.88	0.01	0.88	0.87
Uses hard drugs daily	0.14	-0.04	0.27	0.01	0.66	0.01	0.92	0.43

Continued on following page.

Table A.2 (continued): Baseline statistics and balance test for endline respondents

		Test of randomization balance (N=947)						
	Sample	Assigned therapy		Assigned cash		Assigned both		F-Test
Baseline covariate	Mean	Coeff.	p value	Coeff.	p value	Coeff.	p value	p value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Committed theft/robbery in past 2 weeks	0.54	0.04	0.59	0.02	0.49	-0.03	0.60	0.70
Number of nonviolent stealing incidents	5.07	-0.54	0.50	-0.48	0.67	0.49	0.61	0.80
Number of felony stealing incidents	0.45	0.04	0.85	0.17	0.60	-0.17	0.66	0.85
Disputes and fights in past 2 weeks (0-9)	2.10	0.08	0.88	0.56	0.47	-0.71	0.25	0.56
Aggressive behaviors (mean of 19), z-score	0.00	0.00	0.96	0.11	0.23	-0.17	0.20	0.27
Conscientiousness index (0-24)	15.37	-0.08	0.67	-0.28	0.34	0.07	0.86	0.01
Neuroticism index (0-21)	12.11	-0.09	0.67	0.15	0.65	0.18	0.73	0.36
Grit index (0-21)	13.75	-0.02	0.82	-0.20	0.62	0.14	0.70	0.80
Reward responsiveness index (0-24)	14.72	-0.19	0.40	0.06	0.90	-0.21	0.75	0.61
Locus of control index (0-24)	14.43	-0.11	0.71	-0.54	0.09	0.49	0.26	0.00
Impulsiveness index (0-21)	9.41	0.42	0.38	0.29	0.51	-0.93	0.14	0.41
Self esteem index (0-24)	13.47	-0.11	0.71	-0.11	0.62	0.16	0.69	0.89
Patience in game play index (0-6)	4.10	0.14	0.19	-0.06	0.77	-0.05	0.89	0.61
Time inconsistency in game play index (0-6)	3.28	-0.24	0.04	-0.03	0.78	0.11	0.50	0.01
Risk aversion index (0–3)	1.57	-0.02	0.86	-0.03	0.74	0.10	0.33	0.70
Self-reported patience (mean of 7), z-score	0.00	-0.05	0.75	-0.10	0.32	0.10	0.55	0.42
Declared Risk Appetite (mean of 6), z-score	0.00	0.03	0.88	0.02	0.90	-0.15	0.43	0.86
Cognitive ability (z-score)	0.00	0.13	0.14	0.16	0.18	-0.31	0.02	0.09
Executive function (z-score)	-0.02	0.07	0.21	0.11	0.40	-0.27	0.11	0.25
R-Squared		0.18		0.11		0.34		
p Value on F-statistics on all covariates		0.55		0.77		0.27		

Notes: Column (1) reports the sample mean. A small number of missing values are imputed at the median. Columns (2)-(7) report the coefficients and p values from ordinary least squares regressions of each baseline covariate on three indicators, one for assignment to each treatment arm, controlling for block fixed effects. Column (8) reports the p value from a joint test of statistical significance of all three treatment indicators.

Table A.3: Recruitment neighborhoods

			Estimated males		Recruited		
Phase	Neighborhood	Short description	All	Target	No.	% all	% target
1	Red Light 1	Peri-urban, along the main road from Monrovia to the northeast of the country, residential but the site of one of the major markets in the city, mixed income neighborhood.	23 422	1 171	100	0.4%	8.5%
2	Red Light 2	Peri-urban, along the main road from Monrovia to the NE of the country, residential but the site of one of the major markets in the city, mixed income neighborhood.	36 434	1 822	219	0.6%	12.0%
	Central Monrovia	The area consists of Mamba Point and West Point, one of the busiest business areas in Monrovia.	32 345	1 617	179	0.6%	11.1%
3	New Kru Town	Peri-urban, the north of Bushrod Island, the transit point to counties of the northwest of Liberia, notorious for petty crime.	28 704	1 435	240	0.8%	16.7%
	Logan Town	Peri-urban, the middle of Bushrod Island, next to Freeport of Monrovia, many garages and small shops and booths.	22 100	1 105	86	0.4%	7.8%
	Clara Town	Peri-urban, the south of Bushrod Island, next to Central Monrovia, lots of car-loaders and wheelbarrowers.	23 921	1 196	175	0.7%	14.6%
All			166 926	8 346	999	0.6%	12.0%

Notes: Total male population estimates come from the authors' calculations based on data from Liberia Institute of Statistics and Geo-Information Services (LISGIS). To get an estimated number of target males we assume half are in the age range of 18-35 and take the bottom 10% decile as our targets. Red Light 1 includes Gorbachop, Woodcamp, Reservoir, Pipeline, Soul Clinic, and Sugar Hill. Red Light 2 includes Turtle Base, Chicken Poultry, Ma Kebbeh Gas Station area, Sugar Hill, Bassa Town, Goba Chop Community, Morris' Farm, Bernard Farm, Pipeline Road, Zayzay Community, Coca Cola Factory Community, Plank Field Community, Banana Bush Community, Soul Clinic Community, and Wood Camp. Central Monrovia includes Mamba Point and West Point areas. New Kru Town also includes part of Calwell. Logan town also includes part of Mamba Point and West Point that are not covered in Phase 2.

Tracking to reduce attrition At baseline we were clear about our desire to stay in touch. We took photos and signature samples, and collected as many as ten different ways to contact each respondent. We documented contact information for each respondent, including all the places they said they sometimes stay, plus contact information for the network of people around them who have a more stable location. Respondents were often on the run from the police or other people, and so their contacts might be uncomfortable speaking to enumerators and revealing the respondent's location. Thus, after the baseline survey, we asked respondents to use the enumerator's phone to call their most stable contact and introduce the enumerator and study and give permission.

At each endline, enumerators would typically start with the phone numbers of the various contacts or respondent and try to arrange an appointment. Contacts received no financial incentive. Failing that they would begin visiting the various locations listed. A slight majority of respondents were found within a few hours. In other cases, all leads were cold and more extensive sleuthing and asking around the neighborhood was required. If someone had traveled or moved far away, enumerators either waited until they returned or traveled across the country to find them in person.

On the upper tail, it could take three to four days of physical searching to find the hardest-to-locate people. Enumerators only stopped searching when all possible leads had been exhausted.

Response rates Table A.4 lists survey response rates by treatment group and survey wave (short term, pooling 2- and 5-week surveys, and long term, pooling 11- and 13-month surveys). It also reports the p-value from a t-test of the difference between the response rate in each treatment group and the control group. None of the differences are statistically significant, and all are within about a percentage point of the control group response rate. The control group response rate is a tiny bit lower in the long run surveys and a tiny bit higher in the short run ones. But none of these differences control for covariates or even strata fixed effects, as in the next table.

Correlates of attrition and compliance We analyze the correlates of attrition in Columns 1 and 2 of Table A.5, which reports an OLS regression of an indicator for attrition on selected baseline covariates.[3] There are not significant differences in attrition by treatment group, substantively or statistically. Those who attrit are slightly wealthier and have slightly poorer mental health. In all, the treatment indicators and covariance are jointly significant at $p = 0.047$ so attrition is not ignorable. This is one reason we control for covariates in all treatment effects regressions.

A.4   Treatment compliance

Figure A.1 displays the distribution of class attendance for those assigned to therapy. NEPI did not collect attendance data during the first week (three sessions), so for simplicity we assume that all participants who attended at least one session after week one also attended the first three sessions.

We use two definitions of compliance. Our first measure is defined as "attending at least 8 days of therapy", or about three of the eight weeks. Our second measure is defined as attending at least 80% of sessions (16 classes plus the 3 in the first week).

We analyze the correlates of compliance in Columns 3 through 8 of Table A.5. Being assigned to cash in addition to therapy did not affect the likeliness of attending therapy, which is to be expected since the cash grants were not known to participants until after therapy. The main correlates of

---

[3]We do so to reduce collinearity and thus ease interpretation. Results with full covariates draw similar conclusions.

Table A.4: Survey response rates by wave and treatment status

|	Control	Treatment Only	Cash Only	Treatment + Cash	All
			Treatment group		
Short-term					
# found	384	484	427	433	1728
# unfound	33	48	49	40	170
Response rate	92.1%	91.0%	89.7%	91.5%	91.0%
p-value vs. control		0.65	0.36	0.83	
Long-term					
# found	404	520	474	472	1870
# unfound	36	40	26	26	128
Response rate	91.8%	92.9%	94.8%	94.8%	93.6%
p-value vs. control		0.65	0.18	0.18	
All					
# found	788	1004	901	905	3598
# unfound	69	88	75	66	298
Response rate	91.9%	91.9%	92.3%	93.2%	92.4%
p-value vs. control		1.00	0.84	0.48	

Notes: Survey response rates are calculated as the difference between the total number of respondents at baseline and the number of respondents "unfound" at each endline, all divided by the number of respondents at baseline. Here, "unfound" refers to both respondents we could not locate and those we did locate but who choose to not participate in the survey.

Table A.5: Baseline correlates of survey attrition and treatment compliance for select covariates

	Dependent variable							
	Unfound		Cash Received		Attended >8d of therapy		Attended >19d of therapy	
Baseline covariate	Coeff. (1)	Std. Err. (2)	Coeff. (3)	Std. Err. (4)	Coeff. (5)	Std. Err. (6)	Coeff. (7)	Std. Err. (8)
Assigned to therapy only	-0.006	[.017]						
Assigned to cash only	-0.009	[.018]						
Assigned to therapy & cash	-0.010	[.018]	0.000	[.011]	-0.001	[.027]	-0.019	[.043]
Age	0.001	[.001]	-0.002	[.002]	0.001	[.003]	-0.005	[.005]
Married or living with partner	-0.017	[.020]	-0.007	[.018]	-0.040	[.038]	0.097	[.062]
# children under 15 in household	-0.002	[.002]	0.005	[.002]***	0.003	[.004]	-0.008	[.007]
Years of schooling	0.000	[.003]	0.000	[.002]	0.019	[.005]***	0.013	[.008]
Cognitive skills, z-score	0.004	[.008]	-0.002	[.006]	-0.016	[.017]	0.029	[.027]
Health index (0-6)	0.000	[.005]	-0.001	[.004]	-0.008	[.008]	-0.012	[.015]
Symptoms of depression and distress, z-score	0.011	[.006]*	0.001	[.004]	-0.008	[.012]	-0.062	[.025]**
War experience index (0-14)	-0.004	[.003]*	-0.005	[.002]**	-0.001	[.004]	0.000	[.008]
Summary index of income measures, z-score	0.009	[.007]	-0.004	[.009]	-0.011	[.015]	-0.003	[.024]
Savings stock (USD)	0.000	[0000]	0.000	[0000]	0.000	[0000]	0.000	[0000]
Hours/week working in potentially illicit activities	0.000	[0000]	0.000	[0000]	-0.001	[.001]	-0.002	[.001]*
Hours/week working (total)	0.000	[0000]	0.000	[0000]	0.000	[0000]	0.000	[.001]
Summary index of antisocial behaviors, z-score	0.001	[.009]	0.003	[.008]	0.037	[.015]**	0.020	[.029]
Summary index of self-control skills, z-score	0.012	[.007]*	0.003	[.004]	0.027	[.013]**	-0.006	[.026]
Patience in game play index (0-6)	0.000	[.003]	0.003	[.004]	0.015	[.007]**	0.028	[.011]**
N	999		499		529		529	
Mean of dependent variable	0.078		0.980		0.905		0.637	
R-squared	0.098		0.162		0.299		0.239	
P-value for test of joint significance	0.047		0.620		0.011		0.005	

Notes: Columns (1)–(2) pool all endline survey rounds and report the coefficients and standard errors from an OLS regression of an indicator for attrition on the baseline covariates. Columns (3)–(8) report the coefficients and standard errors from an OLS regression of an indicator for compliance on the same covariates, restricting the samples to people assigned to the respective treatment groups only. block fixed effects are included in all regressions but are omitted from this table. Robust standard errors are clustered at the individual level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure A.1: Distribution of CBT Attendance
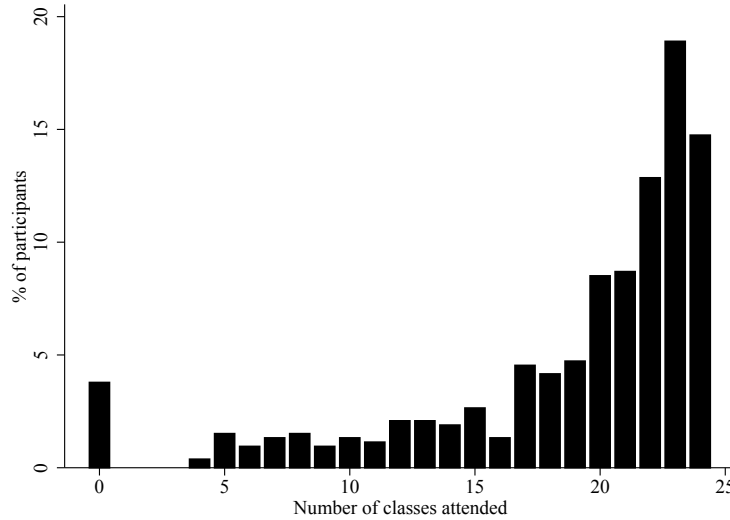
Notes: The figure reports the distribution of therapy attendance. No attendance data was collected during the first week, so we assume for simplicity that all participants who attended at least one session after week one also attended the first three sessions.

compliance in the first three weeks are higher education, higher initial antisocial behaviors, and higher self-control skills. The main correlates of attending at least 80% of the sessions are higher education, better mental health, and patience in game play. Higher initial antisocial behaviors, and higher self-control skills are no longer so relevant.

B Additional intervention details

B.1 Power calculations

After completing the pilot, we decided on a target sample of 1,000. This target was based on maximum program capacity and financial constraints. Based on the pilot, we estimated that the Minimum Detectible Effect for the full 1,000 (with a quarter for each treatment) would be a 0.12 standard deviation change in a standardized dependent variable for a two-tail hypothesis test with statistical significance of 0.05, statistical power of 0.80, an intra-cluster correlation of 0.25, and the proportion of individual variance explained by covariates as 0.10.

B.2 Randomization protocols

For the therapy and cash randomization, men in each block took turns drawing colored chips from an opaque fabric bag. In general, the bag was shaken and then the subject was instructed to turn away and to place one arm into the bag and to draw out a single chip. The color was confirmed and recorded.

In the cash instance, men were randomized in roughly equal sized blocks of about 50 people. Each man was invited into a private room to draw to ensure privacy and safety. This procedure was

explained to the entire group, and all chips were placed into the bag in front of everyone. Then the bag was taken into a private room, and participants were called into the room individually. If they wished, they could inspect the bag to confirm that there were still chips of both colors inside. After everyone present had drawn, staff drew the remaining chips for the no-shows.

In the case of therapy, men were randomized each day, according to how many were recruited and surveyed in that neighborhood. This led to blocks ranging in size from 1 to 20, though the vast majority of blocks contained roughly 7 to 15 people. The draw was not as private as the cash draw, and men observed the outcomes of others drawing at the same time. Those who lost in the therapy randomization were offered a free meal along with the opportunity to discuss their situation with someone, and they were transported to a location of their choosing. A small percentage of the men were visibly upset and refused to engage at this point.

B.3   Therapy

NEPI's standard curriculum tended to be longer and broader than the two noncognitive skill and value changes that we study. For the purposes of this study, we worked with NEPI to streamline and focus the traditional STYL curriculum in two ways. First, we further grounded the approach in terms of CBT, emphasizing more practice over lectures. In general these modifications were quite modest, since the program already incorporated these techniques. Second, we asked NEPI to exclude modules not relevant to their theories of change: interpersonal skills; conflict resolution skills; dealing with war trauma and PTSD; career counseling; and community leadership.

To clarify and validate NEPI's curriculum, a Liberian qualitative researcher acted as a participant observer throughout one of the two Phase 1 pilot classes. Based on NEPI's training materials, our analysis of the theoretical grounding of the therapy, and this participant observation, we and NEPI developed a full program manual for the intervention.[4] The manual details the history and theory of the interventions, guidelines for recruitment of trainers and participants, training suggestions, the full curriculum, and guidelines for out-of–classroom engagement.

Curriculum

The curriculum has eleven main modules, which we present here with some examples of goals and activities:

1. *Transformation.* A tenet of CBT is that the therapist explicitly sets goals with participants and lays out the therapeutic strategy. This module introduces the concept of transformation, its significance, and the processes involved in transforming oneself.

   - The men are introduced to the techniques that will be used (role playing, lectures, storytelling, etc.), homework assignments, home visits, and the reasons for each.
   - The module also introduces ground rules for behavior, in terms of being respectful, practicing listening, waiting your turn, etc. The men do not necessarily have these skills, or haven't exercised them in some time, and learning to abide by these behavioral rules is an important part of the therapy.

---

[4] Available at http://chrisblattman.com/documents/policy/2015.STYL.Program.Manual.pdf.

- Facilitators also begin to teach the songs, slogans, and call-and-response that will be used repeatedly throughout the course. These songs and slogans serve as important reminders of rules of behavior for the men to follow. They also can be used to bring order to a disorderly or inattentive group.
- There are symbolic rituals to indicate a break in their lives. For example, the men write their "street names" and aliases on sheets of paper and they are burned together.

2. *Substance Abuse.* This module defines substance abuse and discusses its ill effects, as well as steps for moving past it. It explicitly encourages participants to reduce their consumption of drugs, alcohol, and tobacco. They are cautioned against cutting drugs entirely, to avoid withdrawal problems.

- Men talk through and list reasons that they use drugs. The idea is to make them consciously aware of the reasons for their own behavior and risk factors in their lives. They also talk through the ill effects. Men talk through publicly about ways in which drugs have adversely impacted their own lives, sharing experiences.
- Men role play situations where they could be pressured to use drugs and practice strategies for saying no.
- An outside speaker comes to the classroom, often a former graduate of the therapy, to talk about their experiences with drugs and what it did to their lives, as well as what strategies they used to emerge. Men discuss strategies they can use in their own lives. They practice some of these as homework and come back to discuss their experiences with the class.

3. *Body Cleanliness.* The module explores the health, psychological, and social benefits of maintaining body cleanliness. Participants are encouraged to change behaviors that alienate them, and to present a public image (such as hair and dress) that promotes positive social interactions with community members.

- Body uncleanliness is defined and highlighted as a problem mainly by getting men to discuss and volunteer their own opinions and experiences in a group.
- The facilitators bring in a hair cutter, an electric shaver, and a set of nail clippers for men to clean up if they like.

4. *Garbage/Dirt Control.* An extension of the previous module, this module highlights the importance of cleanliness in participants' environments, and the ill effects of living in a dirty environment. It aims to help them maintain clean, healthy, and orderly living spaces.

- Facilitators present the men with pictures of dirty and clean homes, businesses, and streets, and men point out different risks and unclean elements, and discuss the consequences.
- Men identify ways they can improve cleanliness where they live (e.g. get a garbage can) and set and execute these plans as homework, to be followed up with home visits.

5. *Anger Management.* This module discusses the causes and effects of anger, the problems with acting out in ways they may later regret. It also provides participants with tools to manage their anger.

- Men discuss the signs and indications of anger, in themselves and others, through discussion and role playing. Facilitators show pictures of angry faces and situations, and men interpret them. The aim is to make them cognizant of these signs.
- Men discuss the causes of anger, and learn to link some of their actions to other people's anger.
- Men discuss and role play the negative consequences of aggression and violence, or share experiences from their own life.
- Men practice nonaggressive responses to angry confrontations in class, such as learning to distract or calm oneself (walking away, doing other activities, starting discussions and de-escalating, or practicing breathing techniques). Men practice these techniques as homework.

6. *Self-Esteem.* This module emphasizes the need for participants to discover themselves in order to begin the path to recovery. This module links their behavioral changes to respect, pride, and confidence.

   - The facilitators try to link poor self-image directly to many of the behaviors they have discouraged in previous modules, both as a cause and consequence.
   - Men discuss ways they can build self esteem, make plans, and execute them as homework.
   - Facilitators work with men to identify worthwhile skills and characteristics they hold that are worthy of others' respect.
   - Men practice shopping for goods in a supermarket or shop as one of the first exposure activities. They work through successes and failures as a group and try again, sometimes with the help of a facilitator.

7. *Planning.* Reviews the steps and components necessary for planning and implementation. The goal of this module is to build participants' capacity to develop short- and long-term plans and understand the processes involved in executing these plans.

   - Planning skills are commonly taught in CBT programs as a method to build new skills. At its most basic, this involves helping the men break down larger plans into smaller steps and helping them work through ways to accomplish those steps, positively reinforcing successes and helping them process challenges and setbacks, often as a group. Men give examples and discuss them together. Another example: Small groups of men are tasked with organizing activities, such as a football match. The larger group listens to the different plans and critiques them.
   - As homework assignments, initially men are tasked with simple tasks (create a short term survival plan for feeding yourself or your family), and then more complex tasks (such as a business plan or home garden).
   - Men are also tasked with identifying a successful friend or family member and determining what steps led to their success. A motivational speaker (usually a past graduate) is also invited to talk about the steps involved in their success and their learnings and setbacks.

8. *Goal Setting.* The module outlines tools participants can use to develop goals, objectives, and indicators for measuring success in their own lives.

- Participants are taught what short and long term goals are (through discussion and examples) and how to set reasonable short- and long-term goals (such as feeding their family, or starting a garden).
- First participants practice setting goals and making plans, and then the larger group discusses and critiques them. Participants then set their own small, short term goals (e.g. changing a behavior, reconciling with a family member, or saving a certain amount this week) and execute these as homework, processing successes and failures as a group.
- Participants discuss the characteristics of good goals (e.g. achievable, measurable, time-bound) and revise goals and plans. They are given poor goals as a group and practice turning them into better goals. Another motivational speaker is used to discuss the role of goal setting in their own life.

9. *Money Business.* This module stresses the importance of engaging in positive spending habits and appropriately managing money. Impulsive spending habits are emphasized. Participants are taught to make plans and prioritize their needs and wants prior to spending their money.

- Men engage in exercises to track their own recent spending to see where their money has gone. They discuss the use and misuse of their own money. As a group they discuss regrets and bad decisions and work through the negative consequences. These are illustrated dramatically through role-playing and skits, followed by discussion.
- Later discussion, role playing and skits focus on techniques for resisting peer pressure and temptation. There is also testimony from a motivational speaker, usually a past graduate of the program.

10. *Money Saving.* The module introduces participants to various saving options and encourages them to reflect on the most suitable saving method for their lives. They practice interactions in informal and formal financial institutions.

- Men discuss the reasons for and advantages of saving and it is explicitly linked to positive self image and esteem in the community. There is another motivational speaker.
- Men learn techniques for saving safely at home without formal institutions. They learn to set and execute saving plans, using their goal setting and planning skills.
- Homework assignments involve saving money they would have otherwise used on things they regret (identified in the previous module). Homework also involves trips to the bank and informal lenders. Prior to these assignments they meet and role play in groups, and their strategies are discussed and critiqued by the larger group. There is also a focus on appropriate presentation and image in these outings.

11. *Challenges and Setbacks.* The module explores potential challenges and setbacks they will face and has them practice positive coping mechanisms needed to effectively overcome them. Challenges and setbacks are framed as a test of one's maturity, potential, and abilities, and an opportunity for improvement.

A note on the approach

Note that in the United States, cognitive behavioral approaches to reducing violence are conscious of the fact that the values and behaviors it encourages could be maladaptive in some situations,

since being violent can also protect people. As a result, these therapies teach people to judge when and where to use aggression.[5] NEPI, in designing the STYL therapy, did not consider the need for educating men on such contingent, adaptive behavior. Rather, their philosophy was that fighting back or retaliating in this context would lead to cycles of violence and an escalation of future risk, not a decrease. NEPI also emphasized how it was also important for the men who passed through STYL to demonstrate to the community that they were not aggressors or violent, to maintain the new image, and retaliation could be counter-productive there.

B.4 Cash grants

We contracted the international non-profit Global Communities (GC) to conduct the registration and cash distribution, as well as oversee NEPI's financial management and implementation schedule. We did so for several reasons:

1. To keep the therapy and the research teams distinct from cash distribution;

2. To coordinate registration and implementation of the two activities;

3. To relieve the research team of project and financial management of the interventions; and

4. To make the intervention as close as possible to a real-world, replicable intervention by other non-profit or state organizations.

For safety, GC developed a highly structured system of cash distribution. GC staff held cash in a car that moved around the neighborhood, to avoid theft. A lottery team with the men gave grant winners a voucher, and put them on a motorbike taxi that was then directed to the street corner where the car with the cash awaited. They were told to approach the car (which had an identifying mark such as a red bag on the dash), hand over their voucher, and receive their cash. The car would then move to a new corner, whose location would be relayed by mobile phone, and the process would repeat.

Anyone who was assigned to the cash treatment but was not present on the day of disbursal was still eligible for the grant. GC attempted to locate them for up to three weeks afterward, and generally succeeded.

C   Formal theoretical model

Our model is rooted in previous models of occupational choice with self-employment (Fafchamps et al., 2014; Udry, 2010; Blattman et al., 2014), but adapted to have a criminal sector as in the broad class of models described by Draca and Machin (2015). We employed a similar model in Blattman and Annan (2015).

---

[5]For instance, in a rough neighborhood, the optimal approach in terms of aggression is to retaliate when provoked, but to avoid starting a fight if not. Therapy aims to help people slow down their reactions and recognize when their automatic response (such as aggression) is and is not appropriate.

C.1 Setup

We model an individual's choice between legitimate business and illicit activities under different conditions—with and without time inconsistency, and with and without financial market imperfections—and assess the predictions for a number of common labor market and crime-reducing interventions: greater punishment, increasing productivity in legitimate business (e.g. through technology or skills improvement), cash or capital transfers, and interventions that shape preferences—either time preferences or personal preferences against illegal behavior.

We use L^b and L^c to denote time spent in legitimate activities (such as petty business) and illegitimate activities (such as crime). Legitimate business produces revenue according to production function $F(\theta, L^b_t, K_t)$, where θ is productivity or individual ability and K is accumulated capital used in business. A person's decision to participate in illegal activity is motivated by the potential gains and costs from such activity. Gains include the expected illegitimate payoff per hour spent in illegal activities, w. Costs include the possibility of apprehension and conviction, which occurs with probability, ρ, and implies a penalty, fL^c_{t-1}. Thus the penalty for criminal behavior is a linear function of hours spent in criminal activities in the previous period[6] The individual's total expected earnings from legitimate and illegitimate activities are $y_t \equiv F(\theta, L^b_t, K_t) + w_t L^c_t - \rho f L^c_{t-1}$. . In addition to investing in business, the individual can also invest or borrow through a riskless asset with constant returns $1 + r$. At each period t, the individual decides how much to invest for next period a_{t+1} and reaps interests ra_t from last period's investments.

Individuals have utility function $U(c, l, \sigma L^c)$, where c denotes consumption and l denotes time for leisure. We also allow for individuals to have direct disutility from engaging in crime, as measured by σL^c, where $\sigma > 0$ implies that implies that illicit work induces some internal penalty such as shame, though in principle it could also reflect social penalties such as a loss of esteem or exclusion from peers and other social networks. We make the standard assumption that $U'_c \geq 0$, $U'_l \geq 0$, $U'_{\sigma L^c} \leq 0$, $U''_{cc} < 0$, $U''_{ll} < 0$, $\partial^2 U / \partial L^2_c \leq 0$ and $F'_\theta \geq 0$, $F'_L \geq 0$, $F'_K \geq 0$, $F''_{\theta\theta} < 0$, $F''_{LL} < 0$, $F''_{KK} < 0$, and $F''_{\theta L} \geq 0$, $F''_{\theta K} \geq 0$, $F''_{LK} \geq 0$.[7] We allow for the individual to have quasi-hyperbolic (β, δ) preferences.

We first consider the case without any uncertainty. The individual's problem is:

$$\max_{c_t > 0, 0 \leq l_t \leq \bar{L}, L^b_t, L^c_t, K_{t+1}, a_{t+1}} U(c_t, l_t, \sigma L^c_t) + \beta \sum_{i=1}^{\infty} \delta^i U(c_{t+i}, l_{t+i}, \sigma L^c_{t+i})$$

$$s.t. \quad c_t + a_{t+1} + K_{t+1} = F(\theta, L^b_t, K_t) + w_t L^c_t - \rho f L^c_{t-1} + (1 + r)a_t \quad for\ each\ t$$

$$a_0 \quad given$$

where $L^b_t + L^c_t + l_t \equiv \bar{L}$.

---

[6]One reason for this modeling choice is because we want to explore the role that quasi-hyperbolic preferences play in the decision to commit crimes when the punishment is in the future not the present.

[7]For ease of analysis, we also assume that the marginal return to capital is infinity for the first unit of capital invested in business, and that as long as there is positive capital input, marginal product of labor for the first unit of labor will be infinity, i.e. $\lim_{K \downarrow 0} F'_K(\theta, L^b, K) = +\infty$ for all L^b and $\lim_{L^b \downarrow 0} F'_K(\theta, L^b, K) = +\infty$ as long as $K > 0$. This assumption guarantees that investments and hours in business will always be positive.

C.2 Occupational choice (and interventions) among time consistent individuals

Without credit constraints

Without time inconsistency ($\beta = 1$) or credit constraints, the set of optimality conditions are:

$$\frac{U'_l(t)}{U'_c(t)} = F'_{L^b}(t) \qquad if \ L^b_t > 0 \tag{1}$$

$$\frac{U'_l(t)}{U'_c(t)} - \sigma \frac{U'_{\sigma L^m}(t)}{U'_c(t)} = w_t - \frac{\rho f}{1+r} \qquad if \ L^m_t > 0 \tag{2}$$

$$1 + r = F'_K(t+1) \qquad if \ K_{t+1} > 0 \tag{3}$$

$$\frac{U'_c(t)}{U'_c(t+1)} = \delta(1+r) \tag{4}$$

$$c_t + a_{t+1} + K_{t+1} = F(\theta, L^b_t, K_t) + w_t L^c_t - \rho f L^c_{t-1} + (1+r)a_t \tag{5}$$

where for ease of notation, we use $U(t)$ to denote $U(c_t, l_t, \sigma L^c_t)$ and $F(t)$ to denote $F(\theta, L^b_t, K_t)$. Since we modeled crime punishment as a potential reduction in future wages, the risk neutral individual will view crime as an occupation with a discounted wage $w_t - \frac{\rho f}{1+r}$.

To find the marginal conditions for engaging in each sector, we first consider the case where illicit activity is not feasible. This would arise naturally if the probability of apprehension is high enough and punishment is heavy enough that $w \ll \frac{\rho f}{1+r}$. In this case the decision to engage in business depends on productivity θ, wealth level and the returns on other financial assets r. We use c^{ba}, L^{ba} and K^{ba} to denote consumption, labor and capital level in this scenario. Each period t, the individual chooses L^{ba}_t to satisfy $\frac{U'_l(c^{ba}_t, \bar{L} - L^{ba}_t, 0)}{U'_c(c^{ba}_t, \bar{L} - L^{ba}_t, 0)} = F'_{L^b}(\theta, L^{ba}_t, K^{ba}_t)$ taking K^{ba}_t as given, and he chooses capital investment K^{ba}_{t+1} to satisfy $F'_K(\theta, L^{ba}_{t+1}, K^{ba}_t) = 1 + r$, taking expected L^{ba}_{t+1} as given.

Now, taking levels of c^{ba}, L^{ba} and K^{ba} as given, we then look at individuals' decision to engage in crime. Individuals will engage in illicit activities if and only if:

$$w_t - \frac{\rho f}{1+r} \geq \frac{U'_l(c^{ba}_t, \bar{L} - L^{ba}_t, 0)}{U'_c(c^{ba}_t, \bar{L} - L^{ba}_t, 0)} + \sigma \frac{-U'_{\sigma L^m}(c^{ba}_t, \bar{L} - L^{ba}_t, 0)}{U'_c(c^{ba}_t, \bar{L} - L^{ba}_t, 0)} \tag{6}$$

which says expected returns from crime are higher than the highest possible marginal rate of substitution between leisure and consumption the individual can achieve without engaging in crime. Since $-U'_{\sigma L^m}/U'_c > 0$, a rise in σ means more people will drop out of crime.

If condition (6) is satisfied and if $K_t > 0$, the individual then chooses L^b_t and L^c_t such that the marginal product of labor in business equals his expected marginal gains from crime, which also equals his marginal rate of substitution between leisure and consumption: i.e. conditions (1) and (2) will be satisfied. Notice L^c_t may not always be positive. The individual will not engage in crime if any or all three of the following happens: w_t is very low relative to the probability of apprehension ρ and punishment f; productivity in business θ is very high; the degree of aversion to crime σ is very high.

Capital investment and hours in business will satisfy condition (3). Notice that w, ρ and f will not affect returns to investment in business.

Interventions that increases the disutility of crime or the size or probability of punishment will reduce time devoted to in crime, but will have no effects on returns in business.[8] However, interventions that increase business productivity θ will not only induce more investment in business, but also reduce involvement in crime. In other words, $\frac{\partial L^c}{\partial \sigma} < 0$, $\frac{\partial L^b}{\partial \sigma}$ is ambiguous, $\frac{\partial L^c}{\partial \theta} < 0$ and $\frac{\partial L^b}{\partial \theta} > 0$. Finally, interventions that provide capital or liquid financial assets, such as a cash windfall, will not affect occupational choice at all, since the individual will already be working at his optimal level in both sectors. The windfall will simply be consumed and saved.

With credit constraints

In this section we consider the model with a simple credit constraint in the form of $a_t \geq 0$–individuals are unable to borrow in any period. We focus our attention on individuals whose initial a_0 is low enough that at some point in his life, the credit constraint is binding. Credit constraints will affect optimal conditions (2) and (3). The optimal condition for capital investment (3) becomes

$$F'_K(\theta, L^b_t, K_t) = max\{1 + r, \frac{1}{\delta}\} \qquad if \ K_{t+1} > 0$$

and the optimal condition for hours in crime (2) becomes

$$\frac{U'_l(t)}{U'_c(t)} - \sigma \frac{U'_{\sigma L^m}(t)}{U'_c(t)} = w_t - \frac{\rho f}{max\{1 + r, \frac{1}{\delta}\}} \qquad if \ L^m_t > 0$$

Notice that $max\{1 + r, \frac{1}{\delta}\} \geq 1 + r$ and $w_t - \frac{\rho f}{max\{1 + r, \frac{1}{\delta}\}} \geq w_t - \frac{\rho f}{1 + r}$. For the impatient individuals whose $\frac{1}{\delta} > 1 + r$, their optimal level of capital investment will be lower than the baseline case because of the credit constraint. They are also have a higher expected returns from crime than in the baseline case, because the low level of business investment also forces them to put a higher discount rate on potential future punishment from crime.

Critical condition (6) becomes

$$w_t - \frac{\rho f}{max\{1 + r, \frac{1}{\delta}\}} \geq \frac{U'_l(c^{ba}_t, \bar{L} - L^{ba}_t, 0)}{U'_c(c^{ba}_t, \bar{L} - L^{ba}_t, 0)} + \sigma \frac{-U'_{\sigma L^m}(c^{ba}_t, \bar{L} - L^{ba}_t, 0)}{U'_c(c^{ba}_t, \bar{L} - L^{ba}_t, 0)}$$

Credit constraints induce more individuals who would otherwise not engage in crime to commit crime. For the impatient individuals, credit constraints increase their hours in crime and reduce their capital investments and hours in business activities.

Interventions that ease the credit constraint, including cash windfalls, will induce more investment in business and reduce involvement in crime. As in the baseline case, $\frac{\partial L^c}{\partial \sigma} < 0$, $\frac{\partial L^b}{\partial \sigma}$ is ambiguous, $\frac{\partial L^c}{\partial \theta} < 0$ and $\frac{\partial L^b}{\partial \theta} > 0$; however, the magnitude the effects of a change in σ or θ will be greater than in the baseline case; the magnitudes also increases with the degree of impatience: $\frac{|\partial L^c / \partial \sigma|}{d\delta} < 0$, $\frac{|\partial L^c / \partial \theta|}{d\delta} < 0$ and $\frac{|\partial L^b / \partial \theta|}{d\delta} < 0$ (notice that the lower the value of σ, the more impatient the individual).

---

[8]The level of investment in business may change depending on the shape of the utility and production functions, but the returns to investment will not change.

C.3   Occupational choice (and the effects of interventions) under time inconsistency

Without credit constraints

Time-inconsistent individuals ($\beta < 1$) will be more reckless in the present. Intuitively, the smaller is β, the more individuals want to enjoy higher consumption today at the expense of future consumption, which means they will borrow more, save less, invest less in business and/or involve more in criminal activities. However, as long as there is a perfect financial market, no one will change their business or criminal activities in order to consume more today–they will simply borrow more (or save less) today through the financial market.

In terms of optimal conditions, in the absence of any credit constraint, the only condition that changes is equation (4), which becomes

$$\frac{U'_c(c_t, l_t, \sigma L_t^c)}{U'_c(c_{t+1}^P, l_{t+1}^P, \sigma L_{t+1}^c)} = [\frac{\partial c_{t+1}}{\partial W_{t+1}}\beta\delta + (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}})\delta] \cdot (1 + r)$$

where W_t denotes total wealth at time t, c_{t+1}^P denotes the individual's predicted future decision about c_{t+1} at time t. For the sophisticates $c_{t+1}^P = c_{t+1}$ while for the naifs $c_{t+1}^P > c_{t+1}$. Compared with the baseline case, the discount factor δ is replaced by the effective discount factor $\frac{\partial c_{t+1}}{\partial W_{t+1}}\beta\delta + (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}})\delta$, a weighted average of the short-run and long-run discount factors $\beta\delta$ and δ where the weights are the next period marginal propensity to consume out of total wealth.

Notice that neither condition (2) nor condition (3) changes, as long as we have no credit constraints. Compared with the baseline, time inconsistency alone will not affect criminal activities or business investment. It would only change the level of savings or debts.

In this case, interventions that aim to correct time consistency will have no effects on either business investment or criminal activities, but will have an effect on consumption, savings and income.

With credit constraints

With credit constraints, in addition to equation (4), optimal conditions (2) and (3) will change as well. Let $\Delta = \frac{\partial c_{t+1}}{\partial W_{t+1}}\beta\delta + (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}})\delta$ be the effective discount factor under (β, δ) preferences and $\tau = \frac{U'_c(c_t, l_t, \sigma L_t^c)}{U'_c(c_{t+1}^P, l_{t+1}^P, \sigma L_t^c)} \cdot \frac{1}{\Delta}$, where c_{t+1}^P denotes the individual's predicted future decision about c_{t+1} at time t. With credit constraints, the Euler equation (4) becomes

$$\tau \geq 1 + r \qquad \text{with equality iff } a_{t+1} > 0$$

and conditions (2) and (3) become

$$\frac{U'_l(t)}{U'_c(t)} - \sigma\frac{U'_{\sigma L^m}(t)}{U'_c(t)} = w_t - \frac{\rho f}{\tau} \qquad \text{if } L_t^m > 0$$

and

$$F'_K(\theta, L_t^b, K_t) = \tau \qquad \text{if } K_{t+1} > 0$$

In addition, critical condition (6) will change accordingly, with $1 + r$ replaced by τ.

Compared with the baseline case, $\tau > 1+r$ as long as an individual is credit constrained (i.e. has no savings). The level of τ will be higher for the sophisticates than for the naifs. However, regardless of their level of sophistication (i.e. the way individuals set their expectations for their future behavior), we know for sure that $\tau > \frac{1}{\delta}$, and the smaller β is (i.e. the more time inconsistent), the higher τ will be.

Compared to the time-consistent credit constrained case, fewer individuals will invest in business, more individuals will engage in crime, business investment levels will be lower, and hours in crime will be higher for everyone. The difference increases with the level of inconsistency (i.e. decreases with β).

Interventions that improve time consistency will shift people away from crime towards business. So will increasing the disutility of crime (though, as in the case without time inconsistency, while $\frac{\partial L^c}{\partial \sigma} < 0$, $\frac{\partial L^b}{\partial \sigma}$ is ambiguous). Increasing business productivity will have similar effects as before: $\frac{\partial L^c}{\partial \theta} < 0$ and $\frac{\partial L^b}{\partial \theta} > 0$. In all of these cases, however, the magnitudes the effects of a change in σ or θ will be greater than under time consistency, and the magnitudes also increase with the both degree of impatience and the degree of time inconsistency: $\frac{|\partial L^c/\partial \sigma|}{d\beta} < 0$, $\frac{|\partial L^c/\partial \theta|}{d\beta} < 0$, $\frac{|\partial L^b/\partial \theta|}{d\beta} < 0$, $\frac{|\partial L^c/\partial \sigma|}{d\delta} < 0$, $\frac{|\partial L^c/\partial \theta|}{d\delta} < 0$ and $\frac{|\partial L^b/\partial \theta|}{d\delta} < 0$. Notice that the lower the value of β, the more time inconsistent the individual is, and similarly, the lower the value of σ, the more impatient the individual is.

C.4 Introducing uncertainty and risk aversion

Three potential sources of risk are uncertainties in business productivity θ, wages from criminal activities w, and the potential punishment after apprehension f. We assume that decisions on business investment and hours in both sectors are made before risks are realized, and that θ, w and f follow independent stochastic processes.

With uncertainties in both the business and illicit sector, business investment and hours in both sectors depend on the variance of returns in both sectors and the level of initial wealth a_0. If both sectors are sufficiently risky, then those with high levels of wealth a_0 will turn away from both activities by reducing K, L^b and L^c and investing instead in other riskless assets. K, L^b and L^c will all be lower than the cases without risk. Those with low levels of initial wealth will not be able to live off savings alone, so they will have to invest more in either or both sectors, depending on the relative riskiness of the two sectors. As long as both sectors are similarly risky, K, L^b and L^c will all be higher; otherwise, if one of the sectors is less risky than the other, individuals will invest more time in that sector. $\frac{L^c}{L^b+L^c}$ will be lower than in the case without uncertainty if returns to crime are more volatile than business returns. One special case would be if individuals face a significantly positive chance of death after committing any crime. This is the equivalent of saying $f = +\infty$ with strictly positive chances. In this case hours in crime will be reduced to zero as long as the probability of apprehension is positive, $\rho > 0$.

With the presence of risk, inventions in θ will have greater effects, because an increase in θ now also makes business *relatively* less risky. A rise in σ will also have a bigger effect than without uncertainty, because risk aversion will reinforce the rise in aversion and further reduce hours in crime.

D Measurement

In this section, we discuss measurement decisions in more detail and report control group means and treatment effects on all of the survey questions that enter an index in the main tables. We do not discuss the economic variables from Table 4 because all index components are displayed in that table. A note of caution: the standard errors have not been adjusted for multiple hypothesis testing, and so patterns across treatment effects within an index are suggestive only.

D.1 Antisocial behaviors

Table D.1 displays treatment effects for all components of our antisocial behaviors index.[9] Section 6.1 of the paper described the construction of these variables in brief. We are not aware of existing scales or measurement tools for Liberia, or even similar populations in sub-Saharan Africa or other low income countries. Thus, in general, our variables grew out of months of field work, qualitative interviews, and survey pre-testing by the authors and their research assistants, in order to understand common offenses and behaviors. Liberians speak a pidgin English and street youth have a slang of their own, and so even where we began with common scales (such as aggressive behaviors) the wording had to undergo extensive translation and testing to make sense. We also added new aggressive behaviors common to the study population and Liberian culture.

D.2 Self control

Table D.2 displays control means and long-term treatment effects for all subcomponents and survey questions in our self-control index, including impulsiveness, conscientiousness, GRIT, and reward responsiveness.

Because all personality questions were selected from questionnaires used in the United States, we first translated them into Liberian English by the enumerators, the authors and their research assistants then pre-tested the questions with young men from the same population as the youth in our study (but not members of the study sample).

These existing scales typically have many more questions than we could use in the survey (or are commonly used in any assessment). These questions are typically organized into sub-scales to capture subcategories of behavior. We selected questions to use based mainly on whether they were easily understood and familiar to pre-test respondents, but we took care to ensure roughly equal proportions of questions from each sub-scale remained.

To ensure that the questions continued to assess the original underlying constructs, we performed two checks. First, within the pre-test data we ensured that groups of questions were correlated or anti-correlated as one would expect given the underlying personality measure (e.g., impulsivity was negatively correlated with conscientiousness). Second, we performed a confirmatory factor analyses to ensure that within scales, questions were answered similarly.

D.3 Time preferences

Table D.3 displays control means and long-term treatment effects for all subcomponents of our forward-looking time preferences index. The summary index consists of eight equally-weighted

---

[9]To save space, we only display long-term (12-13 month) results.

Table D.1: Program impacts on antisocial behaviors, 12–13 month survey only

	Control	ITT regression (N=947)					
		Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of antisocial behaviors, z-score	0.032	-0.083	[.093]	0.132	[.097]	-0.247	[.088]***
Usually sells drugs	0.135	-0.034	[.029]	0.035	[.030]	-0.059	[.029]**
# of thefts/robberies in past two weeks	1.839	0.073	[.395]	0.352	[.388]	-0.728	[.363]**
"Took something behind someone not for you" (stole)	0.275	-0.006	[.067]	0.106	[.072]	-0.063	[.069]
"Corrected someone's mistake" (stole unwatched items)	0.338	-0.067	[.077]	0.091	[.078]	-0.087	[.084]
"Scraped from others" (Cheating)	0.299	-0.037	[.071]	0.023	[.074]	-0.104	[.077]
Pick-pocketed someone	0.094	-0.039	[.039]	0.041	[.041]	-0.074	[.035]**
"Scammed someone" (Sold false goods or conned)	0.118	-0.005	[.038]	-0.044	[.039]	-0.093	[.036]**
"Black deed business" (Con artistry)	0.598	0.104	[.141]	0.016	[.128]	-0.196	[.137]
Mugged someone	0.086	0.075	[.077]	0.033	[.057]	-0.078	[.050]
Armed robbery	0.032	0.001	[.035]	0.021	[.036]	-0.026	[.026]
Disputes and fights in past two weeks, z-score	-0.060	-0.026	[.091]	0.100	[.090]	-0.100	[.077]
Small palava (dispute) with a neighbor	0.152	-0.005	[.062]	0.062	[.059]	0.047	[.065]
Small palava (dispute) with a leader	0.059	0.008	[.049]	0.005	[.041]	-0.008	[.035]
Small palava (dispute) with the police	0.152	0.033	[.064]	0.039	[.053]	-0.050	[.052]
Large fight with a neighbor	0.076	0.021	[.043]	0.023	[.033]	-0.036	[.032]
Large fight with leader	0.100	-0.027	[.047]	0.027	[.045]	-0.084	[.037]**
Large fight with police	0.027	0.021	[.027]	0.045	[.026]*	0.007	[.036]
Physical fight	0.115	-0.004	[.055]	0.087	[.065]	-0.079	[.045]*
Engaged in a fight with a weapon	0.083	-0.028	[.041]	0.023	[.042]	-0.048	[.037]
Fined for a fight	0.051	-0.081	[.062]	0.020	[.069]	-0.040	[.059]
Carries a weapon (typically knife)[†]	0.148	-0.059	[.031]*	0.043	[.035]	-0.066	[.033]**
Arrested in past two weeks	0.118	-0.006	[.024]	0.007	[.025]	-0.033	[.024]
Aggressive behaviors, z-score	0.188	-0.153	[.110]	-0.043	[.107]	-0.339	[.109]***
In the last 4 weeks, have you been quick to react against others?	0.611	-0.026	[.075]	0.047	[.078]	-0.117	[.078]
In the last 4 weeks, have you refused to take advice?	0.596	-0.113	[.081]	-0.042	[.082]	-0.183	[.082]**
Do you sometimes make hard jokes about people?	1.236	-0.013	[.093]	-0.115	[.094]	-0.156	[.092]*
In the last 4 weeks, have you intentionally destroyed property?	0.365	-0.026	[.073]	-0.057	[.072]	-0.110	[.074]
Do you sometimes cheat or scrape from people?	0.547	-0.047	[.084]	0.027	[.081]	-0.073	[.081]
In the last 4 weeks, have you ever had confusion with people about things?	0.606	-0.052	[.076]	0.016	[.079]	-0.123	[.082]

Continued on following page.

Table D.1 (continued): Program impacts on antisocial behaviors, 12–13 month survey only

Outcome	Control mean	ITT regression (N=947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
In the last 4 weeks, did you let others see your frustration when you were frustrated?	0.555	-0.027	[.080]	0.015	[.080]	-0.085	[.082]
In the last 4 weeks, have you threatened other people?	0.360	-0.042	[.071]	-0.027	[.069]	-0.137	[.072]*
In the last 4 weeks, have you taken things from behind other people without asking them?	0.421	-0.047	[.075]	0.074	[.079]	-0.159	[.076]**
In the last 4 weeks, have you easily controlled your vexation when vexed? (-)	1.010	-0.051	[.090]	0.099	[.095]	-0.031	[.098]
Do you get vexed when you lose a game?	1.704	-0.163	[.094]*	-0.242	[.096]**	-0.251	[.095]***
Can you feel fine when you hit or yell at somebody?	0.828	-0.158	[.091]*	-0.051	[.092]	-0.269	[.092]***
If you are under attack can you hit that person to defend yourself?	1.867	-0.010	[.093]	-0.020	[.091]	-0.051	[.094]
When someone teases you, does that make you vexed?	1.493	-0.128	[.097]	-0.013	[.098]	-0.144	[.100]
Do you ever fight to show that you are the stronger person?	0.833	-0.104	[.094]	-0.011	[.094]	-0.153	[.098]
Do you ever damage things as a joke or for fun?	0.680	-0.162	[.084]*	0.010	[.083]	-0.250	[.084]***
Do you ever hurt the person you are playing football with for you to win?	0.882	-0.174	[.091]*	-0.177	[.090]*	-0.141	[.100]
Do you ever use force on somebody to do something for you?	0.567	0.041	[.086]	0.016	[.081]	-0.105	[.086]
Do you ever cuss somebody to do something for you?	0.596	0.054	[.084]	0.066	[.086]	-0.110	[.082]
Verbal/physical abuse of partner, z-score[†]	-0.071	0.142	[.100]	0.233	[.113]**	0.059	[.104]
Last month, did you accuse your woman for getting boyfriend?	0.388	0.115	[.083]	0.120	[.087]	-0.030	[.084]
Last month, did you ever tell your woman you will beat her?	0.381	0.084	[.078]	0.095	[.082]	0.027	[.079]
Last month, did you ever cuss your woman?	0.218	0.023	[.064]	0.123	[.075]	0.036	[.071]
Last month did you push, hit, slap or throw something at your wife or girlfriend?	0.152	0.076	[.056]	0.151	[.065]**	0.087	[.057]

Notes: The table reports long-run (12–13 month) intent to treat estimates of antisocial behavior outcomes. We calculate the impact of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1 [†]These variables were not collected during every phase/round, so their regressions have a smaller sample size.

Table D.2: Program impacts on self control skills, 12–13 month survey only

		ITT regression (N=943)					
	Control	Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of self-control skills	-0.070	0.159	[.090]*	-0.025	[.095]	0.244	[.095]**
Impulsiveness	0.082	-0.178	[.096]*	0.006	[.098]	-0.212	[.099]**
I buy things quick without thinking	1.246	-0.097	[.089]	0.012	[.091]	-0.082	[.092]
I can take action before thinking (-)	0.931	-0.025	[.091]	0.140	[.092]	-0.082	[.090]
I can just talk without thinking	0.744	-0.072	[.082]	0.021	[.086]	-0.208	[.082]**
I am not set or relax at lectures	1.153	-0.124	[.095]	-0.035	[.097]	-0.129	[.099]
I can catch hard time thinking	1.271	-0.074	[.093]	0.014	[.096]	-0.275	[.095]***
I believe in the present rather than the future	1.650	-0.150	[.105]	-0.190	[.106]*	-0.063	[.112]
I able to control myself (-)	0.596	-0.051	[.071]	0.066	[.078]	0.020	[.078]
I spend money on things and regret it later	1.848	-0.045	[.073]	0.017	[.074]	0.047	[.075]
Conscientiousness	0.018	-0.065	[.097]	-0.028	[.100]	0.044	[.097]
I am ready anytime	2.148	-0.267	[.067]***	-0.146	[.066]**	-0.215	[.067]***
I pay attention to things good good	2.305	-0.046	[.050]	-0.054	[.051]	-0.038	[.051]
I get everyday work done right away	2.044	0.006	[.062]	0.037	[.063]	0.108	[.065]*
I make plans and go by them	2.177	0.120	[.057]**	-0.011	[.057]	0.054	[.059]
I catch hard time to do my work (-)	1.704	-0.076	[.070]	-0.059	[.071]	-0.030	[.073]
I do unasked additional work after finishing my work.	1.468	-0.002	[.069]	0.075	[.072]	0.046	[.073]
I can't complete/finish things (-)	1.970	0.033	[.056]	0.026	[.060]	0.054	[.057]
I run away from work (-)	1.788	0.079	[.063]	0.066	[.062]	0.125	[.063]**
Perseverance/GRIT	-0.037	0.116	[.099]	0.057	[.099]	0.105	[.103]
I have overcome hard times to subdue an important challenge	2.020	-0.047	[.063]	-0.009	[.063]	0.060	[.068]
I can think big about my future	2.300	0.079	[.053]	0.094	[.053]*	0.114	[.054]**
Difficult conditions don't discourage me	1.980	0.026	[.065]	-0.090	[.068]	0.007	[.067]
My greatest prayer is to be successful in life	2.468	0.054	[.051]	0.100	[.052]*	0.026	[.055]
I'm trying hard to make it	2.305	0.009	[.053]	-0.008	[.053]	0.015	[.051]
I sometime make a plan but later on change my plan to a different one (-)	1.271	0.085	[.064]	-0.058	[.063]	-0.068	[.066]
I do not think too much about big things/success (-)	1.340	0.025	[.075]	0.086	[.075]	0.055	[.075]
Reward responsiveness	0.072	-0.165	[.102]	0.084	[.100]	-0.242	[.102]**
When I want something I can go to any corner to make sure I get it.	1.473	-0.149	[.075]**	-0.039	[.076]	-0.170	[.076]**
When I go after something, even the devil in hell can't stop me.	1.404	-0.072	[.078]	0.050	[.076]	-0.072	[.080]

Continued on following page.

Table D.2 (continued): Program impacts on self-control skills, 12–13 month survey only

Outcome	Control mean	ITT regression (N=943responses from 943 subjects)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Most of the time, I will do things for no other reason than that I will enjoy them.	1.379	-0.138	[.059]**	-0.006	[.064]	-0.075	[.066]
When I am doing well at something, I will always like to be doing it.	2.059	0.061	[.053]	0.086	[.052]*	0.013	[.056]
When I get something I want, I can jump with happiness and it gives me plenty strength.	2.079	-0.027	[.056]	0.108	[.056]*	-0.096	[.060]
When I see a chance to get something I really like, I can jump with happiness on the spot.	1.966	-0.044	[.065]	0.013	[.065]	-0.129	[.064]**
When good things happen to me, it affects me strongly.	1.611	-0.023	[.069]	-0.082	[.071]	-0.057	[.071]
I can jump with happiness when I win a lucky ticket.	2.207	-0.053	[.057]	0.098	[.053]*	-0.070	[.059]

Notes: The table reports long-run (12–13 month) intent to treat estimates of self control outcomes. N=943 because 4 respondents did not answer all questions. We calculate the impact of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

components: four measures for patience (δ) and four measures for time inconsistency (β). Components come from incentivized game play, hypothetical trade-offs over time, and survey measures. This section reviews the measures in detail and discusses how results change with different index construction and components.

We have four types of measures, and each one yield a proxy of patience and time inconsistency

1. Incentivized trade-offs

   Following the survey, subjects were asked to play a set of "real money games" where they had to make a series of intertemporal choices between money at one point in time versus more money later in time, with some probability of a payout. The average payout was about $3, roughly a day's wages.[10] The first choice was between money now and more money in two weeks; second between two weeks and four weeks; and finally one more question for each of these pairs of delays, but with the numbers modified depending on their first answer (i.e. if they chose to wait, then they were asked again but with a lower reward in the future). This bifurcating design allowed us to glean as much information as possible about their preferences with as few questions as possible, and we pretested the potential payouts to maximize the variance in responses.

   Based on game play, we assigned present and future patience scores for each respondent, ranging from 0 (less patient) to 3 (more patient).[11] We then used the sum of patience scores from the games to put people into 7 increasingly patient bins (0–6), and the difference of scores to put people into 7 increasingly time inconsistent bins.

2. Hypothetical trade-offs

   During the survey questionnaire, well before the incentivized games, we asked respondents to make the exact same series of tradeoffs as above, but in a purely hypothetical setting. We constructed the patience and time inconsistency proxies in exactly the same manner. Our aim was largely methodological, as we were interested in whether people responded differently when games were incentivized rather than hypothetical. This analysis—comparing the consistency and comparability of time preferences over different measures and over time—will be the subject of future methodological work, based on similar data we have collected across several countries and populations.[12] In the meantime, we merely use all available time preference

---

[10]Subjects were told that one of the questions across the next few activities would be picked for payout, and their choice implemented, so that they should pay careful attention to their decisions. We told subjects that if one of the inter-temporal tasks was chosen for payout, and if their individual choice implicated a delayed reward, that we would come back and find them at the appointed time, in their own environment, to pay them.

Since we were typically returning in a few weeks to interview them again, and had interviewed them several times before, this was a reasonably credible commitment. Nonetheless, it could lead us to conflate patience with trust that the survey team would return. By the endline stage (their fifth survey with us), respondents knew us fairly well and knew that we were able to track them (and that we had paid them everything we had promised them in the past).

In fact, for logistical reasons, we also made one of the games a choice between a certain payout now and a lottery between a high and low payout (i.e. a risk preference question) and we selected this risk game for payout with very high probability, such that the intertemporal games were almost never paid out Although we did not technically lie at any point (since we did not mention the probabilities that each task would be paid out) this could be construed as minor deception. None of the respondents brought this up, even after having gone through the process five times.

[11]For example, if a respondent preferred 150 Liberian dollars (or LD, where 1 USD = 60 LD at the time) in a week over 50LD now, and 100 LD in a week over 50 LD now, they received a 3 for their present patience score. If they preferred 50 LD in two weeks over 150 LD in three weeks, and 50 LW in two weeks over 300 LD in three weeks, they received a 0 for their future patience score.

[12]In the meantime, we can see that the means similar (3.96 for the incentivized game versus 3.35 for the hypothetical), but this 15% difference is statistically significant at the 99% level.

measures in our summary index, in the interest of reporting all survey measures used from each family.

3. Hypothetical discount rate

   We also attempted to measure the discount rate in a second way (again, mainly for the methodological study mentioned above). As in Holt and Laury (2002), we asked respondents a series of hypothetical inter-temporal choices for larger amounts of money (on the order of US$10-30, about a week's wages). This was organized as two lists of 11 binary decisions, with a fixed amount right now versus a varying amount in two weeks (or two weeks versus four weeks for the second list). The delayed amount started as strictly less than the sooner amount (e.g. 1000 LD now or 900 in the future), then equal to, and then larger and larger until it was four times as big (1000 LD now or 4000 LD in future).

   We calculated discount rates based on each respondent's first switch from a present preference to a future preference.[13] Those who preferred 900 LD in the future over 1000 LD in the present received a discount rate of .9, while those who always preferred money earlier received a discount rate of 4. We then took the average of the inverse of the present (now versus 2 weeks) and future (in 2 weeks versus 4 weeks) discount rate as our measure of patience, and the difference between future and present as our measure of time inconsistency.

4. Self-reported survey questions

   We asked respondents six qualitative questions to gauge their self-reported levels of patience and time inconsistency.[14] For example, respondents were asked to place themselves on a ladder from 0 (least patient) to 5 (most patient) as one measure of self-reported patience, and how much they agree with statements such as "When I get money, I spend it quickly" as a proxy of time inconsistency. Specific questions are displayed in Table D.3.

By reporting all measures collected in the endline survey, three-quarters of our time preference measures are hypothetical rather than based on incentivized games. For robustness purposes, in Table D.3 we also report a summary index of the incentivized games only.

D.4   Anti-criminal and anti-violent values

Table D.4 displays control means and long-term treatment effects for survey questions in the anti-criminal and anti-violent values index.

D.5   Other outcomes

Table D.5 displays the control group means and long-term impacts of all the survey questions underlying the variables in Table 8, including prosocial behavior, post-traumatic stress, personality traits, quality of appearance, substance abuse, quality of social networks, well being, and executive function. We describe each of these measures in Section 6.6 of the main paper.

---

[13]Enumerators continued down the list, and (oddly) a nontrivial fraction switched multiple times. We use the first switch only. Furthermore, about 17% of respondents preferred less money in the future as a commitment device, especially if they were expecting a large purchase coming soon.

[14]Dohmen et al. (2011) and Jamison and Karlan (2011)show that basic self-reported attitudes on risk and time preferences can be externally valid.

Table D.3: Program impacts on time preferences, 12–13 month survey only

Outcome	Control mean	ITT regression (N= 947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of forward-looking time preferences	-0.149	0.149	[.102]	0.105	[.102]	0.209	[.105]**
Patience (δ) summary index, z-score	-0.240	0.170	[.103]*	0.145	[.096]	0.258	[.099]***
Incentivized trade-offs (0-6)	3.941	0.033	[.173]	0.266	[.166]	0.163	[.172]
Hypothetical trade-off (0-6)s	3.345	0.295	[.222]	0.376	[.221]*	0.331	[.222]
Hypothetical discount rate (.9 to 4)	2.238	-0.177	[.091]*	-0.121	[.091]	-0.251	[.091]***
Self-reported survey questions on patience, z-score	-0.363	0.091	[.093]	-0.032	[.092]	0.139	[.092]
Placing oneself on a 5-rung patience ladder (0-5)	3.757	0.003	[.109]	-0.048	[.107]	0.146	[.110]
"I consider myself a patient person." (0-3 scale)	2.267	0.091	[.063]	0.043	[.063]	0.129	[.065]**
"If I make good money, I save some for future problems." (0-3 scale)	2.108	0.041	[.072]	-0.078	[.072]	-0.035	[.073]
Time inconsistency (β) summary index, z-score	0.129	-0.072	[.083]	0.018	[.087]	-0.059	[.084]
Incentivized trade-offs (-3 - 3)	0.227	-0.024	[.071]	-0.010	[.074]	-0.038	[.073]
Hypothetical trade-offs (-3 - 3)	0.192	0.039	[.099]	-0.042	[.100]	-0.078	[.101]
Hypothetical discount rate (-3.1 - 3.1)	0.021	0.048	[.068]	0.060	[.072]	0.117	[.070]*
Self-reported (mean of 3), z-score	-0.473	0.185	[.100]*	-0.020	[.096]	0.136	[.103]
"If I get money, I spend it quickly." (0–3)	2.618	0.047	[.077]	0.020	[.075]	0.121	[.076]
"If I make good money, I spend a lot celebrating with friends." (0–3)	2.956	0.106	[.077]	-0.036	[.075]	0.070	[.079]
"I avoid going around friends who waste money." (0–3)	1.902	0.141	[.079]*	-0.025	[.079]	0.036	[.082]
Incentivized trade-offs (patience and time inconsistency), z-score	-0.047	0.029	[.081]	0.101	[.081]	0.084	[.084]

Notes: The table reports long-run (12-13 month) intent to treat estimates of outcomes that were not a priori specified as of primary interest. We calculate the impact of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. The final variable in column 1 is the average of the patience and time inconsistency measures from our incentivized trade-offs. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table D.4: Program impacts on anticriminal and antiviolent values, 12–13 month survey only

Outcome	Control mean	ITT regression (N=947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of anticriminal and antiviolent values	0.070	-0.087	[.101]	-0.005	[.102]	-0.203	[.099]**
Attitudes toward use of violence (11 questions)	0.051	0.074	[.120]	0.029	[.122]	0.001	[.119]
If a stranger in the community robs one of your neighbors, is it okay for your neighbor to send people to abuse the stranger?	0.127	0.032	[.039]	-0.037	[.038]	0.003	[.040]
If a man owes you money but refuses to pay, would it be all right to take something from his home?	0.036	-0.009	[.026]	0.002	[.028]	-0.016	[.026]
If someone kills a known criminal in your community, it is ok if the police don't investigate?	0.122	0.022	[.035]	0.058	[.041]	-0.015	[.034]
If a rogue steals from a market, and the storekeepers chase him and beat him, should the storekeepers be punished? (-)	0.447	0.100	[.058]*	0.043	[.059]	0.052	[.058]
What if they mistakenly kill the rogue? Should the storekeepers be punished? (-)	0.189	0.011	[.045]	0.063	[.050]	0.038	[.048]
Suppose your friend's wife ran off with another man and stole his money and belongings. Would it be good for him to chase and beat that woman?	0.138	0.093	[.041]**	0.048	[.042]	0.020	[.042]
If a 'highman' is caught should his property be destroyed?	0.434	-0.043	[.057]	-0.036	[.059]	-0.002	[.059]
Suppose someone tries to steal your friend's girlfriend. Your friend is talking about getting the boys together to threaten the man. Would you join in?	0.107	-0.025	[.034]	0.005	[.035]	-0.035	[.034]
If a local leader is corrupt and taking money for himself, is it all right if he is beaten by the community?	0.163	-0.052	[.042]	-0.065	[.045]	-0.062	[.045]
If a wife challenges her husband in public, is it ok for her husband to beat her?	0.179	0.002	[.043]	-0.036	[.044]	-0.027	[.045]
Suppose a man rapes a girl in your community. He is arrested, but he bribes the police and goes free. If some people beat that police man, would you join in?	0.173	0.002	[.043]	-0.001	[.047]	0.037	[.046]
Attitudes toward criminality (12 questions)	0.044	-0.109	[.118]	-0.106	[.117]	-0.287	[.116]**
Imagine that a Chinese man cheats another man of his wage for carrying a load. Would it be ok for that man to take his friends to beat the Chinese man hard, and to take his money?	0.081	-0.049	[.026]*	-0.044	[.029]	-0.070	[.028]**

Continued on following page.

Table D.4 (continued): Program impacts on self-control skills, 12–13 month survey only

	Control	ITT regression (N=947)					
		Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
If your best friend has some counterfeit money that need to be washed with mercury, will you join him in the search of the mercury to get lots of money?	0.168	0.009	[.042]	-0.002	[.045]	-0.052	[.042]
Is it wrong for someone in your community to make money by entering someone's house at night and taking their valuable things?	0.137	-0.070	[.038]*	-0.087	[.035]**	-0.042	[.041]
After being down on luck, a man finds a wallet in a taxi cab with lots of money and the driver's license. Should he take the money for himself before giving the wallet to the police or a radio station?	0.365	-0.063	[.054]	-0.021	[.058]	-0.098	[.057]*
Your family has no money to pay rent and will soon be homeless. You see $100 USD fall from a man's back pocket. Is it ok to take it and not tell him?	0.310	-0.022	[.053]	-0.019	[.053]	-0.048	[.054]
Is it okay for a man to hook up his house to an electrity cable from his family member to take free electricity behind his back, even though his family member will have to pay for it?	0.157	-0.026	[.044]	0.024	[.043]	-0.054	[.045]
A man's wife was in labour pain but he had no money. He came across a market woman's wallet filled with all her day's earnings. Is it okay for him to take the bag?	0.360	-0.009	[.055]	-0.058	[.054]	-0.113	[.055]**
If a man's $100 USD bill is hanging out of his back pocket and you have a clear change to jerk it with no one catching you, will you feel bad if you take it, even if you don't need it?	0.198	0.004	[.044]	0.001	[.044]	-0.067	[.043]
If a stranger left his room door open with all his valuables in there and no one is around, is it okay to correct his mistake?	0.107	-0.022	[.036]	-0.034	[.034]	-0.056	[.035]
Would you feel fine if your hustle was selling diazepan, bubble, 10-10, or any other drugs llike this?	0.137	-0.025	[.038]	-0.017	[.040]	-0.046	[.039]
If a friend left his room door open with his money inside and no one is around, is it okay to take his money?	0.091	-0.042	[.031]	-0.030	[.031]	-0.070	[.030]**
If you are working in a business, and you have access to your boss's money, would you feel fine taking that money for yourself, if you plan to return it later?	0.244	0.026	[.049]	0.005	[.051]	-0.046	[.051]
	XXX						

Table D.4 (continued): Program impacts on self-control skills, 12–13 month survey only

		ITT regression (N=947)					
	Control	Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Attitudes toward political violence (6 questions)	0.096	-0.143	[.118]	-0.002	[.123]	-0.225	[.119]*
Suppose the politician that you support does not win and there was no cheating. If that politician asks you to protest for him, would you do so?	0.129	-0.020	[.035]	-0.012	[.037]	-0.055	[.035]
Suppose the politician that you support does not win and there was no cheating. If that politician asks you to loot for him, would you do so?	0.079	-0.015	[.029]	0.009	[.032]	-0.042	[.029]
Suppose the politician that you support does not win and there was no cheating. If that politician asks you to loot for him for 200 LD, would you do so?	0.080	-0.011	[.031]	0.028	[.033]	-0.024	[.031]
Suppose the politician that you support does not win but there was cheating. If that politician asks you to protest for him, would you do so?	0.204	-0.075	[.042]*	-0.047	[.044]	-0.085	[.044]*
Suppose the politician that you support does not win but there was cheating. If that politician asks you to loot for him, would you do so?	0.151	-0.057	[.038]	-0.008	[.039]	-0.072	[.039]*
Suppose the politician that you support does not win but there was cheating. If that politician asks you to loot for him for 200 LD, would you do so?	0.130	-0.034	[.037]	0.024	[.037]	-0.056	[.037]

Notes: The table reports long-run (12–13 month) intent to treat estimates of outcomes that were not a priori specified as of primary interest. We calculate the impact of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

In order to measure executive function, our behavioral protocol included three interactive activities drawn from economics and psychology.[15]

Planning behaviors We used a series of mazes to test planning behavior. Mazes were unknown to nearly all respondents. Subjects were shown an example maze on paper and then given 2, 2, and 3 minutes respectively to complete increasingly difficult mazes. Each had two entry points, one of which almost immediately led to a dead end. The main outcome of the mazes was the subject's ability to pause and plan their approach before completing the maze (i.e. did they plan their approach before choosing a starting point). As outcomes, we measure "time to first touch", or the amount of time spent planning prior to engaging in the maze; and number of mistakes (or "backtracks") in Maze 3, the hardest maze, which required the most planning and by which time participants had learned the concept of the maze. On average subjects took 18 seconds to plan for Maze 3 (SD = 23 seconds).

Behavioral inhibition and cognitive flexibility We developed the "arrows game", a modified directional Stroop task, a class of tasks that assess inhibitory control. Here subjects were shown a sequence of large black or white arrows that pointed either up or down and were first told to respond "up" or "down" to each arrow ("arrows baseline"). In the second version they were again shown the arrows but now were told to state the opposite direction; this constitutes producing the less common response while suppressing the more common response and is an assessment of inhibition ("arrows inhibition"). Finally, in a third version subjects were told to switch between two approaches: if the arrow was white they were to state the actual direction, but the opposite direction if the arrow was black. This is commonly called 'switching' and is an assessment of cognitive flexibility, the ability to move rapidly between two goals as the situation demands ("arrows switching"). For each version, the outcome data included total time to completion and the number of correct/incorrect responses out of 32 arrows. On average subjects made .33 errors (SD = 1.5) on arrows baseline, 2.4 errors (SD = 3.5) on arrows inhibition, and 3.9 errors (SD = 3.9) on arrows switching. Arrows took on average 25 seconds (SD = 17.7), 38 seconds (SD = 45.8), and 46 seconds (SD = 28.7) for baseline, inhibition, and switching separately.

Working memory Working memory is the ability to hold something in mind when it is no longer present in the environment and then manipulate it. The digit span task is an assessment of working memory. The digit span tasks involved the enumerator saying a random sequence of digits (1-9) out loud with a short pause between each digit, followed by the respondent repeating them back either in the same (forward-digits) or the reverse (backwards-digits) order. The enumerator began by giving two 2-digit numbers (one at a time) and recording the responses. If the subject correctly reported either of the numbers back, the enumerator would do the same with 3-digit numbers, and so on up to a maximum of 9 digits. As soon as the subject incorrectly reported both examples at a given level or span the enumerator moved on to the next activity (backwards-digits). The reverse digit span was done the same way, except that the subject was instructed to repeat the digits in

---

[15]Across all behavioral tests administration was standardized. First, a clinical psychologist and economist trained enumerators in test administration. Next, in collaboration with experienced enumerators and research assistants, a comprehensive protocol was developed and used by all future enumerators. Enumerators were also instructed to answer clarifying questions and were taught the over-arching concept within each game so they could address questions/alleviate concerns without straying from the central concepts of the tests. This tight control over the testing situation allowed us to collect relatively sophisticated measures of cognitive function and behavioral responses to rewards in a constrained and otherwise under-resourced testing environment.

the opposite order that the enumerator gave them (e.g., "three, zero, one") On average subjects were able to remember 5.5 digits forward (SD = 1.23) and 3.33 digits backwards (SD = 1.03). Each activity existed as two slight variants (e.g. changing the numbers in the gambles). These activities were alternated in the 2 versus 5-week endlines and the 12 versus 13-month endlines, so that participants were never asked identical questions too close together in time.

D.6 Distinguishing between different measures of "self-control"

Our summary indexes distinguish between self-control skills (assessed by various psychological scales), economic time preferences (using incentivized and hypothetical games), and (as an "other" outcome) executive function. Here we discuss the decision to separate these measures and what happens when we relax that assumption.

First, we treat the difference between time preferences and self control skills as an empirical question. As reported in Section 6.4, they are positively and significantly correlated but with a correlation of 0.33 it is unclear whether they are distinct or not. As we report in Table D.6, combining both into an equally-weighted index leads to large increases in the measure for both the therapy-only group (0.16 SD in the short term, 0.18 SD in the long term) and therapy and cash group (0.22 SD in the short term, 0.26 SD in the long term).

Second, we separate executive function from self control as well.

A main reason is that these abilities mature over the lifespan, and psychologists and neuroscientists have emphasized the importance of early-stage investments over late-stage investments because the neuroscientific principle of developmental plasticity, and data from randomizing young children into different early investments suggests that early, but not later investments shape cognitive function (Nelson, 2007).

This is not to say that they are not highly correlated or have common roots early in life. A large literature documents that in some extreme populations (e.g., individuals with substance abuse disorder, kids with ADHD) many of these indices of 'self control' co-vary. That is, kids with ADHD have deficits in performance on inhibition tasks (e.g.,Barkley (1997)). These same children, by definition, behave impulsively and appear to be more sensation or risk seeking. Taken together, many have taken this covariance as evidence that these traits are interdependent. There is even a small neuro-imaging literature which suggests that these different forms of impulsivity are subserved by the same neural areas (Aron (2007)).

Nonetheless, there are many hints in the psychology and neuroscience literature that this is an oversimplification. For example, even within extreme populations, sensation seeking and impulsivity, measured similarly, may be differentially linked with behavior (Ersche et al. (2010)). In typical developing children, successfully resisting temptation on delay of gratification tasks is not predicted by performance on inhibitory control tasks, but the strategies employed in attempting to resist temptation is (Eigsti et al. (2006)).

In fact, the best test is to do what we have done here: randomly assign individuals to an intervention which shifts one of these indices and observe if they all move together. The fact that we see no improvement in executive function is consistent with the skills being different. In Table D.6 we test the combined measures formally, and we do not observe significant increases in a measure combining self control with executive function. Furthermore, their correlation is only 0.15, less than half of the correlation between self control and time preferences

Table D.5: Impacts on other outcomes

				ITT regression (N=947)			
	Control	Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Prosocial behavior	0.018	0.041	[.088]	-0.075	[.085]	-0.017	[.090]
Number of active groups	0.692	0.062	[.085]	-0.157	[.083]*	-0.070	[.087]
Group leader index (4 questions)	0.049	-0.018	[.107]	-0.068	[.106]	-0.023	[.112]
Community leader now (0-1)	0.030	-0.003	[.019]	0.012	[.019]	0.003	[.019]
Big man in community (0-1)	0.108	0.002	[.034]	0.001	[.033]	0.017	[.035]
Group community leader (0-1)	0.193	-0.006	[.039]	-0.025	[.039]	-0.010	[.040]
Start group (0-1)	0.189	-0.012	[.040]	-0.050	[.041]	-0.036	[.040]
# of public good contributions (6 months)	2.500	3.352	[3.685]	0.484	[3.585]	4.092	[5.474]
Trust index (4 questions)	-0.079	-0.001	[.145]	0.075	[.142]	0.039	[.141]
Trusts relatives (0–3)	1.990	-0.171	[.158]	-0.091	[.158]	-0.100	[.154]
Trusts leaders (0–3)	0.796	0.001	[.140]	-0.005	[.136]	-0.126	[.140]
Trusts NGOs (0–3)	1.686	0.035	[.163]	0.051	[.160]	0.060	[.162]
Trusts IPA (0–3)	2.330	0.165	[.132]	0.280	[.125]**	0.255	[.132]*
Post-traumatic stress (5 questions)†	0.136	-0.124	[.101]	-0.061	[.100]	-0.167	[.104]
Sit and think bad things (0–3)	1.401	-0.108	[.095]	0.002	[.096]	-0.214	[.097]**
Have bad dreams (0–3)	1.119	-0.219	[.090]**	-0.028	[.089]	-0.106	[.091]
Seems like bad things happening again (0–3)	0.842	-0.108	[.088]	-0.023	[.089]	-0.098	[.092]
Sweat when thinking about bad things (0–3)	1.000	0.000	[.091]	-0.033	[.086]	-0.178	[.091]*
Stay away from places(0–3)	1.372	0.045	[.095]	-0.064	[.096]	0.037	[.098]
Neuroticism (8 questions)†	-0.019	0.044	[.097]	0.035	[.102]	-0.153	[.096]
Can feel sad one time (0–3)	1.951	0.044	[.049]	0.077	[.050]	-0.039	[.056]
Worry about things (0–3)	1.606	-0.049	[.067]	0.041	[.070]	-0.093	[.069]
Quick to feel threatened (0–3)	1.325	0.033	[.066]	-0.063	[.071]	-0.050	[.068]
Easily offended (0–3)	1.542	0.092	[.067]	0.051	[.072]	0.052	[.073]
Easily stressed (0–3)	1.340	-0.011	[.066]	-0.036	[.071]	-0.067	[.071]
Easily disturbed (0–3)	1.330	-0.013	[.064]	-0.056	[.068]	-0.076	[.065]
Feel relaxed most of time (-) (0–3)	1.355	-0.080	[.072]	-0.082	[.071]	-0.160	[.071]**
Nothing can bother me (-) (0–3)	1.034	0.092	[.062]	0.152	[.066]**	0.062	[.067]
Locus of control (8 questions)†	0.010	-0.032	[.101]	-0.111	[.098]	-0.022	[.106]
Your choices determine your future (0–3)	2.123	0.104	[.054]*	0.130	[.057]**	0.063	[.055]
You have small control over your life (-) (0–3)	1.163	-0.088	[.058]	-0.119	[.060]**	-0.105	[.061]*
Success in business is due to luck (-) (0–3)	1.300	-0.024	[.070]	-0.098	[.070]	0.060	[.078]
Trying hard can make your life better (0–3)	2.340	0.005	[.050]	-0.002	[.050]	-0.046	[.050]
Your can bring your plans into fruition (0–3)	2.123	0.076	[.048]	-0.039	[.045]	0.039	[.046]
Bad things in life are due to bad luck (-) (0–3)	1.616	-0.091	[.072]	-0.115	[.073]	-0.069	[.077]
People are homeless because of their own fault (0–3)	1.384	-0.011	[.074]	0.010	[.078]	-0.029	[.080]
Success comes from hard work (0–3)	2.379	-0.038	[.055]	0.003	[.059]	0.041	[.058]
Self esteem (8 questions)†	-0.071	0.078	[.098]	0.060	[.100]	0.190	[.101]*
Satisfied with yourself? (0–3)	1.872	0.119	[.069]*	0.017	[.072]	0.063	[.073]
Feel useless? (-) (0–3)	1.591	0.076	[.074]	-0.059	[.080]	0.137	[.077]*
Think everything will fail? (-) (0–3)	1.438	0.071	[.065]	0.082	[.066]	0.119	[.067]*

Continued on following page.

Table D.5 (continued): Program impacts on other outcomes, 12–13 month survey only

Outcome	Control mean	ITT regression (N=947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Don't get enough respect? (-) (0–3)	1.768	0.058	[.070]	0.071	[.073]	0.064	[.074]
Feel at least as good as most people (0–3)	1.985	-0.044	[.060]	0.045	[.059]	-0.049	[.066]
Feel like a good person but doing nothing (-) (0–3)	1.837	-0.052	[.067]	0.067	[.069]	-0.018	[.070]
Do business as well as most people (0–3)	2.202	-0.024	[.053]	-0.028	[.055]	0.011	[.053]
Feel shame (-) (0–3)	1.202	-0.010	[.073]	-0.047	[.078]	0.144	[.076]*
Quality of appearance (6 questions)	0.016	-0.102	[.078]	-0.085	[.077]	-0.109	[.082]
Condition of clothes (0–2)	1.554	-0.063	[.050]	-0.069	[.051]	-0.037	[.051]
Condition of shoes (0–2)	1.569	-0.017	[.094]	0.005	[.093]	-0.058	[.097]
Cleanliness of breath (0–2)	0.713	-0.006	[.037]	0.004	[.035]	-0.036	[.038]
Cleanliness of face (0–2)	1.608	-0.077	[.047]	-0.077	[.046]*	-0.083	[.051]
Cleanliness of hair (0–2)	1.397	-0.028	[.056]	-0.001	[.055]	-0.026	[.056]
Cleanliness of fingernails (0–2)	1.520	-0.087	[.052]*	-0.090	[.051]*	-0.062	[.055]
Substance abuse (0-3 index)	1.091	-0.065	[.061]	0.063	[.062]	-0.057	[.060]
Usually drinks (0–1)	0.766	-0.063	[.042]	-0.047	[.042]	-0.045	[.043]
Usually uses marijuana (0–1)	0.490	-0.024	[.035]	0.018	[.035]	-0.042	[.036]
Usually takes hard drugs (0–1)	0.196	-0.005	[.031]	0.079	[.033]**	0.006	[.031]
Quality of social networks (z-score)	0.066	0.063	[.092]	-0.044	[.092]	0.139	[.095]
Peers (mean of 20 questions)	0.040	0.011	[.088]	-0.070	[.089]	0.017	[.090]
Friend in school (0–1)	0.802	0.003	[.039]	-0.060	[.043]	0.013	[.041]
Friend participates in community meetings (0–1)	0.624	0.005	[.047]	-0.049	[.049]	0.052	[.048]
Friend goes to church (0–1)	0.756	-0.013	[.043]	-0.064	[.044]	-0.048	[.045]
Friend works hard (0–1)	0.970	-0.077	[.026]***	-0.086	[.026]***	-0.083	[.027]***
Friend has business or job (0–1)	0.660	-0.036	[.049]	-0.075	[.050]	-0.052	[.051]
Friend saves money regularly (0–1)	0.645	-0.051	[.048]	-0.017	[.049]	-0.043	[.050]
Friend gives good advice (0–1)	0.883	-0.027	[.033]	-0.070	[.035]**	-0.060	[.036]*
Friend likely to share (0–1)	0.827	0.032	[.037]	0.036	[.037]	0.033	[.038]
Friend cheers you up (0–1)	0.817	-0.045	[.041]	-0.002	[.040]	-0.053	[.041]
Fried trusted to guard valuables (0–1)	0.751	0.014	[.043]	-0.027	[.044]	-0.001	[.044]
Friend begs for money (-) (0–1)	0.772	-0.014	[.041]	0.003	[.042]	0.005	[.043]
Friend gets drunk regularly (-) (0–1)	0.787	0.002	[.042]	0.027	[.041]	0.032	[.042]
Friend uses drugs regularly (-) (0–1)	0.706	0.011	[.043]	-0.060	[.044]	0.015	[.043]
Friend pickpockets regularly (-) (0–1)	0.692	0.011	[.037]	-0.019	[.038]	0.027	[.038]
Friend burglarizes (-) (0–1)	0.782	0.050	[.039]	0.030	[.039]	0.056	[.040]
Friend is armed robber (-) (0–1)	0.909	0.037	[.028]	0.018	[.027]	0.064	[.026]**
Friend gambles (-) (0–1)	0.660	0.006	[.047]	-0.022	[.047]	0.038	[.047]
Friend is ex-combatant (-) (0–1)	0.548	-0.039	[.044]	-0.044	[.043]	-0.027	[.045]
Friend is former commander (-) (0–1)	0.721	0.015	[.043]	0.015	[.042]	0.064	[.042]
Friend has small fights with others (-) (0–1)	0.772	0.035	[.043]	-0.022	[.043]	0.007	[.044]
Friend has large fights with others (-) (0–1)	0.843	0.039	[.036]	0.030	[.036]	0.019	[.036]

Continued on following page.

Table D.5 (continued): Program impacts on other outcomes, 12–13 month survey only

		ITT regression (N=947)					
Outcome	Control mean	Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Family (4 questions)	-0.019	0.124	[.099]	0.070	[.100]	0.129	[.097]
Sees family often (0–3)	1.904	0.129	[.094]	-0.021	[.096]	0.094	[.097]
Family concerned about you (0–3)	2.061	0.114	[.102]	0.090	[.103]	0.113	[.102]
Receives encouragement from family (0–3)	2.107	0.062	[.104]	0.048	[.102]	0.070	[.103]
Receives help from family when in trouble (0–3)	1.178	0.120	[.106]	0.132	[.104]	0.155	[.103]
Ex-commanders (4 questions)	0.176	0.004	[.076]	0.026	[.078]	-0.139	[.074]*
Friend is ex-military commander (0–1)	0.279	-0.015	[.043]	-0.015	[.042]	-0.064	[.042]
Close relations with ex-military commander (0–1)	0.208	0.018	[.042]	0.038	[.041]	-0.045	[.042]
Receives job from ex-military commander (0–1)	0.046	-0.006	[.021]	0.012	[.024]	-0.024	[.021]
Reports to commander now (0–1)	0.026	0.004	[.016]	-0.005	[.015]	-0.010	[.016]
"Big men" (5 questions)	0.120	0.001	[.155]	-0.130	[.152]	0.071	[.160]
Has patron for job (0–1)	0.238	-0.007	[.042]	0.011	[.045]	0.067	[.045]
Has patron for business needs (0–1)	0.099	0.008	[.030]	-0.042	[.029]	0.001	[.030]
Has patron for food (0–1)	0.347	0.041	[.047]	0.052	[.048]	0.051	[.050]
Has patron for school fees (0–1)	0.079	0.046	[.030]	0.028	[.030]	0.024	[.030]
Has patron for housing (0–1)	0.361	0.120	[.049]**	0.018	[.050]	0.070	[.052]
Summary index of subjective well being (3)†	-0.020	0.057	[.072]	-0.009	[.072]	0.184	[.074]**
Absolute level today, 1-10 ladder	3.655	0.244	[.270]	0.024	[.271]	0.665	[.287]**
Overall level, (1-30)	13.198	0.504	[.711]	0.323	[.723]	1.912	[.746]**
Happiness (1–10)	4.000	0.274	[.285]	0.344	[.302]	0.837	[.295]***
Satisfaction (1–10)	4.315	-0.087	[.279]	0.053	[.294]	0.886	[.303]***
Health (1–10)	4.883	0.317	[.305]	-0.074	[.295]	0.188	[.307]
Relative to your community, (1-40+)	14.980	-0.026	[.821]	0.522	[.804]	2.642	[.859]***
Relative wealth (1–10)	3.035	0.081	[.224]	0.103	[.231]	0.606	[.248]**
Relative respect (1–10)	4.965	-0.034	[.312]	-0.023	[.316]	0.827	[.334]**
Relative power (1–10)	3.649	0.045	[.289]	0.334	[.269]	0.700	[.293]**
Relative access (1–10)	3.348	-0.138	[.249]	0.088	[.265]	0.489	[.256]*
Executive function (z-score)	0.110	-0.094	[.077]	-0.078	[.076]	-0.109	[.078]
Arrow time (z-score)	0.043	-0.034	[.045]	0.007	[.044]	-0.057	[.053]
Game 2	0.077	-0.013	[.010]	0.001	[.009]	-0.020	[.012]
Game 3	0.077	-0.015	[.033]	0.005	[.032]	-0.028	[.037]
Arrow error (z-score)	0.070	0.038	[.077]	-0.002	[.080]	0.045	[.078]
Game 2	0.258	-0.002	[.074]	-0.077	[.081]	0.004	[.076]
Game 3	0.179	0.066	[.078]	0.075	[.080]	0.073	[.076]
Maze (z-score)	0.060	-0.175	[.089]**	-0.116	[.088]	-0.162	[.091]*
First touch	0.010	-0.066	[.071]	-0.143	[.078]*	-0.100	[.073]
Backtrack	0.344	-0.189	[.088]**	-0.026	[.082]	-0.135	[.091]
Back digit	0.034	-0.053	[.071]	-0.060	[.074]	-0.108	[.075]

Notes: The table reports long-run (12-13 month) intent to treat estimates of outcomes that were not a priori specified as of primary interest. We calculate the impact of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1 †These variables were not collected during every phase/round, so

Table D.6: Effects of combining various measures of self control

Outcome	Round	Control mean	ITT regression (N=947)					
			Therapy only		Cash only		Both	
			ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Self control, z-score	2–5w	-0.037	0.085	[.098]	-0.147	[.104]	0.037	[.096]
	12–13m	-0.070	0.159	[.090]*	-0.025	[.095]	0.244	[.095]**
Time preferences, z-score	2–5w	-0.202	0.179	[.098]*	0.071	[.099]	0.318	[.099]***
	12–13m	-0.149	0.149	[.102]	0.105	[.102]	0.209	[.105]**
Executive function, z-score	2–5w	-0.103	0.076	[.075]	0.059	[.077]	0.024	[.085]
	12–13m	0.110	-0.094	[.077]	-0.078	[.076]	-0.109	[.078]
Self control + time preferences	2–5w	-0.066	0.166	[.072]**	0.016	[.072]	0.221	[.073]***
	12–13m	-0.212	0.183	[.088]**	0.060	[.088]	0.255	[.091]***
Self control + executive function	2–5w	-0.095	0.100	[.070]	-0.006	[.072]	0.037	[.077]
	12–13m	0.060	-0.012	[.073]	-0.072	[.074]	0.009	[.075]
Self control + time preferences + executive function	2–5w	-0.038	0.097	[.042]**	0.009	[.042]	0.129	[.043]***
	12–13m	-0.124	0.106	[.051]**	0.035	[.051]	0.149	[.053]***

Notes: The table reports intent to treat estimates of the effect of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. Heterosketastic robust standard errors are reported in brackets. Because there are two endline surveys per round per individual, each surveyed individual enters twice into the regressions, and standard errors are clustered by individual. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

E   Additional treatment effects analysis

E.1   Robustness of treatment effects

Our robustness tests focus on the five main summary outcomes. First, in Table E.1, we show robustness to alternative ways of constructing the indexes and pooling or averaging of endlines. Columns 2–4 report results from the main paper for comparison. Recall that in this main specification we averaged endline surveys (at 2 and 5 weeks, and 11 and 13 months), took an index of composite measures rather than individual survey questions and used equal weights. In columns 5–7, we do the same except use randomization inference to assess statistical significance. In columns 8–10, we pool our composite measures from both endline surveys and cluster our standard errors by individual. In columns 11–13, we do the same except weight each survey question equally. In columns 14–16, we use covariance-weighted indexes from Anderson (2008) and average both endlines.[16]

The conclusions from these three specifications are quantitatively similar to those from the main specification. Exceptions are as follows:

- The impact of cash and therapy on the covariance-weighted antisocial behaviors index is not significant in the long term at conventional levels. This is because half of this index's weights come from domestic violence and number of arrests, two components that were unaffected by treatment. If we exclude domestic violence from the index and recalculate covariance weights, cash and therapy lead to a .26 standard deviation decline in antisocial behaviors in the long run (columns 19, significant at 99% level).

- Cash increases antisocial behaviors in the long term in some specifications. In Column 15 we see that after a year the men who report cash only increased their antisocial behaviors by 0.17 standard deviations. In the other specifications, the coefficients are positive as well but smaller and not statistically significant. One possibility is that receiving a cash grant and failing, or having the money stolen, reinforces men's participation in crime. This is largely speculative, however.

Next, we check for robustness to alternative attrition scenarios by bounding treatment effects. We impute outcome values for unfound individuals at different points of the observed outcome distribution. The most extreme bound, from Manski (1990), imputes the minimum value for unfound treated members and the maximum for unfound controls. Following Karlan et al. (2015), we also calculate less extreme bounds by imputing relatively high values of the dependent variables for missing control group members, and relatively low values for missing treatment group members.[17] Specifically, we impute missing dependent variables for the treatment (control) group as the found treatment (control) mean minus (plus) 0.10, 0.25, or 1 SD of the found treatment (control) distribution. Note these imply large and systematic differences between missing treatment and control members—Columns 8 – 10 assume unfound control group member outcomes are roughly 2 SD greater than unfound treatment group member outcomes.

---

[16]For this index, each component is weighted by the inverse of the covariance matrix of all index components. Outcomes that are highly correlated with each other receive less weight while outcomes that are uncorrelated receive more weight as they represent new information. We cannot covariance weight the pooled endlines, since they are unbalanced in the sense that some outcome measures appear in only one endline while others appear in both.

[17]This assumes the dependent variable points in the positive direction. If treatment leads to a decrease in the outcome variable, as is the case for antisocial behaviors and antiviolent and anticriminal values, we impute in the opposite direction (i.e smaller values for control, larger values for treatment).

Table E.1: Robustness to alternative index construction and outcome measurement

Outcome, z-score	Round	Main specification			Endline averages, RI			Pooled mean effects of composite measures		
		Therapy only	Cash only	Both	Therapy only	Cash only	Both	Therapy only	Cash only	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Antisocial behaviors	2-5w	-0.249 [.088]***	-0.079 [.091]	-0.308 [.089]***	-0.249 [.082]***	-0.079 [.082]	-0.308 [.082]***	-0.210 [.074]***	-0.081 [.074]	-0.263 [.072]***
	12-13m	-0.083 [.093]	0.132 [.097]	-0.247 [.088]***	-0.083 [.083]	0.132 [.083]	-0.247 [.082]***	-0.078 [.081]	0.103 [.082]	-0.209 [.075]***
Income	2-5w	0.233 [.089]***	0.486 [.097]***	0.465 [.090]***	0.233 [.090]***	0.486 [.091]***	0.465 [.090]***	0.161 [.087]*	0.318 [.094]***	0.223 [.087]**
	12-13m	0.063 [.111]	-0.059 [.105]	-0.024 [.103]	0.063 [.092]	-0.059 [.093]	-0.024 [.092]	0.139 [.093]	-0.054 [.090]	0.025 [.089]
Self control	2-5w	0.085 [.098]	-0.147 [.104]	0.037 [.096]	0.085 [.091]	-0.147 [.092]	0.037 [.092]	0.085 [.090]	-0.147 [.096]	0.037 [.088]
	12-13m	0.159 [.090]*	-0.025 [.095]	0.244 [.095]**	0.159 [.085]*	-0.025 [.085]	0.244 [.084]***	0.159 [.084]*	-0.025 [.088]	0.244 [.088]***
Time preferences	2-5w	0.179 [.098]*	0.071 [.099]	0.318 [.099]***	0.179 [.091]*	0.071 [.090]	0.318 [.091]***	0.168 [.069]**	0.091 [.070]	0.276 [.070]***
	12-13m	0.149 [.102]	0.105 [.102]	0.209 [.105]**	0.149 [.093]	0.105 [.093]	0.209 [.093]**	0.163 [.088]*	0.094 [.088]	0.217 [.090]**
Antiviolent and anticriminal values	2-5w	-0.206 [.094]**	-0.187 [.096]*	-0.180 [.097]*	-0.206 [.083]**	-0.187 [.083]**	-0.180 [.083]**	-0.206 [.086]**	-0.187 [.088]**	-0.180 [.090]**
	12-13m	-0.076 [.088]	0.026 [.088]	-0.177 [.086]**	-0.076 [.077]	0.026 [.077]	-0.177 [.077]**	-0.083 [.083]	0.015 [.082]	-0.181 [.081]**
N			947			947			1861	
Index of composite measures?			Y			Y			Y	
Endlines averaged?			Y			Y			N	
Randomization inference?			N			Y			N	
Covariance-weighted index?			N			N			N	

Table E1 (continued): Robustness to alternative index construction and outcome measurement

		Pooled mean effects of survey questions			Covariance-weighted average endlines			Covariance-weighted average endlines (no domestic violence)		
Outcome, z-score	Round	Therapy only	Cash only	Both	Therapy only	Cash only	Both	Therapy only	Cash only	Both
	(1)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Antisocial behaviors	2-5w	-0.192	-0.073	-0.270	-0.168	-0.076	-0.277	-0.176	-0.094	-0.292
		[.071]***	[.071]	[.070]***	[.091]*	[.091]	[.085]***	[.090]**	[.090]	[.083]***
	12-13m	-0.086	0.041	-0.257	-0.008	0.173	-0.175	-0.085	0.100	-0.263
		[.078]	[.078]	[.072]***	[.091]	[.097]*	[.089]*	[.099]	[.101]	[.090]***
Income	2-5w	0.161	0.318	0.223	0.226	0.529	0.447			
		[.087]*	[.094]***	[.087]**	[.084]***	[.091]***	[.078]***			
	12-13m	0.139	-0.054	0.025	-0.014	-0.021	-0.091			
		[.093]	[.090]	[.089]	[.112]	[.104]	[.107]			
Self control	2-5w	0.069	-0.102	0.024	0.099	-0.137	0.051			
		[.066]	[.070]	[.064]	[.099]	[.104]	[.096]			
	12-13m	0.121	-0.029	0.171	0.165	-0.032	0.252			
		[.057]**	[.061]	[.061]***	[.091]*	[.095]	[.097]***			
Time preferences	2-5w	0.149	0.021	0.214	0.120	0.036	0.249			
		[.069]**	[.069]	[.070]***	[.094]	[.095]	[.095]***			
	12-13m	0.165	0.027	0.192	0.132	0.106	0.180			
		[.084]*	[.083]	[.086]**	[.100]	[.101]	[.104]*			
Antiviolent and anticriminal values	2-5w	-0.291	-0.297	-0.299	-0.236	-0.227	-0.205			
		[.135]**	[.137]**	[.139]**	[.110]**	[.113]**	[.114]*			
	12-13m	-0.058	0.006	-0.153	-0.075	0.008	-0.220			
		[.064]	[.063]	[.061]**	[.102]	[.100]	[.099]**			
N		1870			947			947		
Index of composite measures?		N			Y			Y		
Endlines averaged?		N			Y			Y		
Randomization inference?		N			N			N		
Covariance-weighted index?		N			Y			Y		

Notes: The table reports the robustness of our results to alternate index construction and measurement outcome. In columns 2–4, we report results from our main specification, where we average composite measures and do not cluster standard errors. In columns 5-7, we do the same but use randomization inference to get our standard errors. . In columns 8–10, we pool our endline surveys, weight composite measures equally, and do not cluster standard errors. . In columns 11–13, we pool our endline surveys, weight each survey question equally, and cluster standard errors by individual. In columns 14–16, we weight components using a covariance-weight from Anderson (2008) and average both endlines. In columns 17–19, we remove domestic violence from our antisocial behaviors index, weight survey questions using a covariance weight from Anderson (2008), and average both endlines.. *** p<0.01, ** p<0.05, * p<0.1

Table E.2 reports ITT estimates under these attrition scenarios. Our results are generally robust to these alternate specifications. When X = 0.25 SD, we still observe large and statistically significant changes in antisocial behaviors, self control, time preferences, and antiviolent and anticriminal values in both the short- and long-term. When X = 1 SD, our estimates of treatment effects lose significance but generally point in the correct direction. Meanwhile, the Manski bound brings us closer to having no treatment effects in the long term.

E.2   Both versus just one treatment

In this section, we compare the effects of receiving one treatment versus receiving both therapy and cash. Specifically, we test whether the coefficients on either therapy only or cash only in Section 6 are statistically different from the coefficients on therapy and cash. Table E.3 displays the mean difference between treatment effects and corresponding p-value for each of our five main outcome variables.

Our results indicate that cash and therapy compliment each other in reducing antisocial behaviors in the long-run, while therapy compliments cash in improving self control skills and reducing violent attitudes in the long-run.

E.3   Short versus long term treatment effects

In discussing our results, we emphasize differences between outcomes 2–5 weeks after the intervention and outcomes 12–13 months after the intervention. In this section, we test whether short- and long-term impacts are the same. We pool our short-term results with our long-term results and run the following OLS regression:

$$Y_{ij} = \beta_0 + \beta_1 ShortTerm_i + \beta_2 T_i + \beta_3 (ShortTerm_i \times T_i) + \mathbf{X}_i \lambda + \gamma_j + \omega_t + \varepsilon_{ij} \tag{7}$$

where $ShortTerm$ is an indicator for outcomes measured in weeks 2 or 5, and T is an indicator for treatment group assignment. In our application, we have three treatment groups (therapy only, cash only, and therapy and cash), include baseline controls and block fixed effects, and cluster our standard errors at the individual level i. The size and direction of β_3 determine whether the treatment effects we observe in the short-term are the same as those observed in the long-term.

Table E.4 reports these estimates for our five main family indexes. For many outcomes, we cannot reject that β_3 is zero. In particular, the short- versus long-term effects of both therapy and cash are not statistically distinguishable. There are two exceptions worth noting. First, while the cash-only group experienced the largest increase in the composite income measure 2–5 weeks after the intervention, these effects diminished a year later. Second, while all three treatment groups saw decreases in antisocial behaviors in the short term, the effects of cash alone and therapy alone subsided 12–13 months later.

E.4   Crime: Disaggregated and annualized impacts

Table E.5 reports the incidence of specific crimes reported in the two weeks prior to the long run survey, breaking down the total number of crimes into the type of crime reported. For consistency,

Table E.2: Robustness to alternative attrition scenarios

Outcome, z-score	Round	Impute missing dependent variable with mean = (-) X SD for missing control (treatment) individuals									"Worst case"		
		X = .10 SD			X = .25 SD			X = 1 SD			Manski bound		
		Therapy only	Cash only	Both	Therapy only	Cash only	Both	Therapy only	Cash only	Both	Therapy only	Cash only	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Antisocial behaviors	2-5w	-0.222 [.083]***	-0.063 [.084]	-0.292 [.083]***	-0.197 [.083]**	-0.038 [.084]	-0.266 [.084]***	-0.071 [.089]	0.087 [.091]	-0.137 [.090]	0.161 [.119]	0.318 [.123]***	0.109 [.121]
	12-13m	-0.076 [.088]	0.127 [.092]	-0.235 [.084]***	-0.060 [.088]	0.142 [.092]	-0.221 [.084]***	0.018 [.090]	0.219 [.094]**	-0.153 [.086]*	0.195 [.116]*	0.367 [.119]***	-0.005 [.121]
Income	2-5w	0.222 [.083]***	0.460 [.090]***	0.424 [.083]***	0.199 [.083]**	0.437 [.090]***	0.401 [.084]***	0.085 [.088]	0.321 [.095]***	0.285 [.089]***	-0.081 [.103]	0.157 [.110]	0.117 [.105]
	12-13m	0.046 [.105]	-0.054 [.098]	-0.027 [.097]	0.030 [.105]	-0.069 [.098]	-0.041 [.097]	-0.049 [.107]	-0.146 [.100]	-0.109 [.099]	-0.316 [.138]**	-0.405 [.132]***	-0.336 [.127]***
Self control	2-5w	0.063 [.090]	-0.153 [.094]	0.032 [.087]	0.041 [.091]	-0.177 [.095]*	0.008 [.088]	-0.071 [.095]	-0.297 [.099]***	-0.111 [.092]	-0.339 [.121]***	-0.579 [.129]***	-0.408 [.120]***
	12-13m	0.135 [.087]	-0.049 [.090]	0.230 [.091]**	0.120 [.087]	-0.064 [.091]	0.216 [.091]**	0.042 [.090]	-0.141 [.093]	0.147 [.094]	-0.105 [.103]	-0.287 [.106]***	0.018 [.108]
Time preferences	2-5w	0.167 [.091]*	0.073 [.090]	0.307 [.091]***	0.142 [.091]	0.048 [.090]	0.281 [.092]***	0.018 [.096]	-0.076 [.095]	0.153 [.096]	-0.249 [.121]**	-0.342 [.123]***	-0.133 [.124]
	12-13m	[.083]	[.090]	[.091]***	[.083]	[.091]	[.082]***	[.090]	[.093]	[.094]	[.101]	[.100]**	[.105]
Antiviolent and anticriminal values	2-5w	-0.181 [.086]**	-0.144 [.088]	-0.161 [.090]*	-0.161 [.086]*	-0.122 [.088]	-0.139 [.090]	-0.060 [.090]	-0.012 [.093]	-0.029 [.094]	0.157 [.114]	0.230 [.122]*	0.220 [.118]*
	12-13m	-0.075 [.083]	0.021 [.082]	-0.170 [.082]**	-0.061 [.083]	0.034 [.082]	-0.158 [.082]*	0.007 [.085]	0.099 [.084]	-0.100 [.084]	0.135 [.101]	0.215 [.100]**	0.022 [.105]
N		999			999			999			999		

xlii

Notes: The table reports intent to treat estimates of each treatment arm in the short and long term under alternative attrition scenarios. We impute missing dependent variables. In columns 2 – 10, we impute missing dependent variables for the treatment group as the found treatment mean minus a multiple of the standard deviation of the treatment distribution. Similarly, we impute missing dependent variables for the control group as the found control mean plus a multiple of the standard deviation of the control distribution. In columns 11 – 13 we apply Manski bounds, imputing the minimum value for unfound treated members and the maximum for unfound controls. Each regression controls for baseline covariates and neighborhood-phase fixed effects. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heteroskedastic robust standard errors are reported in brackets. . *** p<0.01, ** p<0.05, * p<0.1

Table E.3: Program impacts on therapy only versus therapy plus cash

	Dependent variable (N=947)									
	Antisocial behaviors		Income		Self control skills		Time preferences		Antiviolent and anticriminal values	
	Mean diff.	p-value	Mean diff.	p-value	Mean diff.	p-value	Mean diff.	p-value	Mean diff.	p-value
Independent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Both vs just therapy										
Short term survey	-0.05	0.53	0.17	0.10	0.03	0.78	0.20	0.06	0.03	0.76
Long term survey	-0.22	0.02	-0.16	0.13	0.05	0.60	0.10	0.34	-0.12	0.17
Both vs just cash										
Short term survey	-0.22	0.01	-0.08	0.44	0.21	0.04	0.24	0.03	-0.03	0.77
Long term survey	-0.47	0.00	-0.06	0.55	0.29	0.00	0.13	0.21	-0.20	0.02

Notes: The table reports the mean difference between receiving both treatments and just one from intent to treat estimates, and corresponding p-values. We calculate the impact of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

we shift from the incidence of drug selling reported in Table 3 to the frequency—the number of times men reported selling drugs in the past two weeks.

Control men committed 2.54 crimes in the previous two weeks, and this fell by almost one crime with therapy plus cash. All types of crime decreased by 20 to 100% with cash and therapy, but the statistically significant (and largest proportional) reductions are in burglary, muggings, and scams (e.g. the sale of non-existent goods, or down-payments for a hidden fortune). We do not adjust p-values for multiple hypothesis testing and so these comparisons across crimes should be taken with caution. In general the coefficients are negative and large in proportion to the control mean (20 to 100%) across all types of crime.

If this decline were persisted for the year, it would translate to 26 fewer crimes per person each year. Given the $530 cost of the two interventions, this is roughly $21 per crime, ignoring any other benefits of the program.

E.5 Heterogeneity analysis

Table E.6 reports impact heterogeneity from an OLS regression of the antisocial behaviors summary index on baseline level of antisocial behaviors, treatment indicators, and interactions between treatment and baseline antisocial behaviors, controlling for baseline covariates and block fixed effects.[18]

Therapy decreased the incidence of antisocial behaviors for the average participant, but men exhibiting more antisocial behavior at baseline saw larger declines. For example, men with average

---

[18]Recall that our measure of antisocial behaviors is a standardized index with mean zero. Therefore, the coefficient on the treatment indicator represents the treatment effect for an individual with mean level of antisocial behavior at baseline, while the coefficient on the interaction term is the additional effect for individuals whose baseline level of antisocial behaviors was 1 standard deviation higher than average.

Table E.4: Testing the difference between short- and long-term impacts of each treatment arm

	Antisocial behaviors		Income		Self control skills		Time preferences		Antiviolent and anticriminal values	
Independent variable	β (1)	S.E. (2)	β (3)	S.E. (4)	β (5)	S.E. (6)	β (7)	S.E. (8)	β (9)	S.E. (10)
β_1										
Short term survey	0.132	[.066]**	-0.132	[.083]	0.042	[.089]	0.089	[.064]	0.011	[.076]
β_2										
Therapy only	-0.013	[.076]	0.128	[.094]	0.118	[.084]	0.126	[.089]	-0.084	[.079]
Cash only	0.094	[.080]	0.010	[.091]	-0.064	[.087]	0.050	[.086]	0.007	[.082]
Both	-0.177	[.071]**	0.043	[.088]	0.209	[.085]**	0.269	[.088]***	-0.182	[.076]**
β_3										
Short term X Therapy only	-0.176	[.081]**	0.052	[.120]	0.019	[.117]	0.076	[.088]	-0.094	[.098]
Short term X Cash only	-0.186	[.085]**	0.335	[.113]***	-0.019	[.121]	0.115	[.084]	-0.147	[.100]
Short term X Both	-0.087	[.081]	0.187	[.105]*	-0.114	[.118]	0.076	[.088]	-0.001	[.097]
Observations	3582		1754		1765		3643		2665	

Notes: The table reports the difference between short-term treatment effects and long-term treatment effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. We pool four endline surveys per individual (2 short-term and 2 long-term) and cluster standard errors by individual. *** p<0.01, ** p<0.05, * p<0.1

Table E.5: Impacts on crime incidence, in the last two weeks and annualized extrapolation

	Cash + therapy ITT, 12–13 month endline				Annualized impact	
	Control mean	Coeff.	Std. Err.	% change	Control mean	Cash + therapy
	(1)	(2)	(3)	(4)	(5)	(6)
# crimes, past two weeks	2.54	-0.994	[.438]**	-39%	66.1	-25.8
# times sold drugs, past two weeks	0.70	-0.266	[.188]	-38%	18.3	-6.9
# thefts/robberies, past two weeks	1.84	-0.728	[.363]**	-40%	47.8	-18.9
Selling/switching fake goods	0.27	-0.063	[.069]	-23%	7.1	-1.6
Stealing unwatched items	0.34	-0.087	[.084]	-26%	8.8	-2.3
Overcharging or cheating	0.30	-0.104	[.077]	-35%	7.8	-2.7
Burglary	0.09	-0.074	[.035]**	-79%	2.5	-1.9
Con artistry/scams	0.12	-0.093	[.036]**	-79%	3.1	-2.4
Pickpocketing	0.60	-0.196	[.137]	-33%	15.5	-5.1
Mugging	0.09	-0.078	[.050]	-91%	2.2	-2.0
Armed robbery	0.03	-0.026	[.026]	-82%	0.8	-0.7
Arrested in past two weeks	0.12	-0.033	[.024]	-28%	3.1	-0.8

Notes: Columns (1) to (4) report the same ITT regression as in Table 3, with robust standard errors in brackets. Columns (5) and (6) simply multiply the two week estimates by 26 weeks to generate an estimated annual impact per person.

*** p<0.01, ** p<0.05, * p<0.1

levels of antisocial behaviors at baseline who were assigned to both therapy and cash experienced a 0.19 standard deviation decline in their level of antisocial behaviors 12–13 months later, but men whose initial level of antisocial behaviors was a standard deviation higher than average experienced about double the decline.

Our results also indicate that in the long run, men with high levels of initial antisocial behavior who received a cash grant actually increase their anti-social acts. This is especially interesting given that the effects of cash on occupational choice and income disappeared after a year. One possibility is that this increase in antisocial behavior is a reaction to the failed attempt at legitimate livelihoods, but these results are more speculative than anything else.

E.6 Program impacts on occupational choice

To measure changes in occupational choice, we asked respondents at each endline whether they had engaged in 22 occupations, from farming to petty business, trades, and formal jobs. For each occupation, we collected self-reported earnings and hours in both the last week and the week prior. We use these to calculate the total earnings and hours variables. With two endline surveys, we have four weeks of employment data per person both in the short-term and in the long term.

We can also calculate hours by occupations each week, aggregating our 22 occupations into 5 mutually exclusive categories:

1. Non-agricultural high-school work, which includes trading and office work

2. Non-agricultural low-skill business, which includes selling from a shop, selling at a table, buying and selling, engaging in petty trade, and conducting small business

Table E.6: Impact heterogeneity based on initial levels of antisocial behavior

Outcome	Round (1)	Coefficient on: (2)	Therapy only		Cash only		Assigned to both	
			Coeff. (3)	Std. Err. (4)	Coeff. (5)	Std. Err. (6)	Coeff. (7)	Std. Err. (8)
Antisocial behaviors	2–5w	Treatment indicator	-0.188	[.072]***	-0.088	[.073]	-0.257	[.072]***
		Interaction term	-0.159	[.092]*	-0.056	[.089]	-0.235	[.077]***
	12–13m	Treatment indicator	-0.023	[.078]	0.104	[.078]	-0.190	[.071]***
		Interaction term	-0.037	[.096]	0.209	[.101]**	-0.179	[.075]**

Notes: We regress our family index of antisocial behaviors on baseline level of antisocial behaviors, treatment indicators, and interactions between treatment and baseline antisocial behaviors, controlling for baseline covariates and block fixed effects. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

3. Non-agricultural low-skill wage labor, which includes contract work, carloading, car-washing, peim-peim riding, carrying loads, guarding, housecleaning, and construction

4. Agricultural work, which includes farming and fishing,

5. Illicit work, which includes selling drugs, stealing, gambling, gold rubber, and scavenging.

Table E.7 reports ITT estimates on the average of the two weeks of data. While we generally observe no changes in overall average hours worked per week (the one exception is those assigned to cash only work approximately 15% more hours per week in the short-term), treatment effects how participants allocate their time. In the short-term, all three treatments cause participants to shift from illicit work to non-agricultural low-skill business. Those assigned to both therapy and cash experience the largest decline in illicit work. Time spent in illicit work falls 38% 2–5 weeks after implementation relative to the control group, and is 17% less than the control group one year later (although the latter is not statistically significant). Although the cash-only group more than doubles its weekly hours spent in non-agricultural low-skill business in the short-term, these effects phase out 12–13 months later.

F   Survey data validation details

F.1   Variables

We selected six variables for validation, all with recall periods of two weeks. We chose outcomes with varying degrees of salience (or memorability) and potential social stigma and experimenter bias. The variables were:

1. *Stealing.* The survey asked how many times in the last two weeks the respondent stole someone's belongings or deceived or conned someone of money or goods.[19] Based on our fieldwork, we hypothesized that stealing would be the most salient and least socially desirable of all six measures.

---

[19]The survey also measured more serious forms of theft, such as armed robbery, but our qualitative validation focussed on non-violent theft.

Table E.7: Program impacts on occupational choice

Outcome	Round	Control mean	Therapy only ITT	Therapy only Std. Err.	Cash only ITT	Cash only Std. Err.	Both ITT	Both Std. Err.
			ITT regression (N=947)					
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average hours/week of work, past 2 weeks	2–5w	36.773	1.044	[2.879]	7.439	[2.907]**	1.787	[2.966]
	12–13m	34.273	1.030	[2.550]	0.681	[2.525]	-1.416	[2.493]
Non-agricultural high-skill work	2–5w	1.717	0.028	[.962]	0.512	[.894]	0.431	[1.006]
	12–13m	3.903	-1.539	[1.080]	-1.046	[1.069]	-0.853	[1.145]
Non-agricultural low-skill business	2–5w	8.904	1.607	[1.991]	10.956	[2.339]***	6.214	[2.178]***
	12–13m	8.880	3.465	[1.683]**	1.859	[1.695]	2.960	[1.728]*
Non-agricultural low-skill wage labor	2–5w	16.011	1.537	[2.172]	-2.652	[2.090]	-0.627	[2.200]
	12–13m	15.804	-0.524	[2.242]	-1.684	[2.107]	-2.902	[2.067]
Agricultural work	2–5w	0.700	-0.576	[.401]	-0.501	[.459]	-0.627	[.411]
	12–13m	0.569	-0.293	[.280]	0.060	[.390]	-0.325	[.319]
Illicit work	2–5w	9.281	-3.339	[1.608]**	-2.793	[1.599]*	-3.486	[1.685]**
	12–13m	6.007	0.029	[1.522]	1.590	[1.491]	-0.995	[1.345]

Notes: The table reports intent to treat estimates of the effect of each treatment arm in the short and long run, controlling for baseline covariates and block fixed effects. The income summary index is the standardized mean of three composite outcomes (themselves first standardized). Heterosketastic robust standard errors are reported in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

2. *Gambling.* The survey asked how many times in the last two weeks the respondent gambled or bet on sports. Beforehand, we hypothesized gambling had a lower level of salience and sensitivity than stealing, but was still somewhat stigmatized.

3. *Marijuana use.* The survey asked how many times in the last two weeks the respondent smoked marijuana. Marijuana use is not socially acceptable across Liberian society overall, but is fairly prevalent in our target demographic. We initially hypothesized underreporting could arise not so much from social stigma but from the discouragement of drug use in the therapy treatment.

4. *Homelessness.* The survey asked how many times in the last two weeks the respondent had to sleep outside, on the street, or in a market stall because they had no other place to sleep or stay. This is a salient variable where we hypothesized respondents might have under-reported from embarrassment or over-reported in order to appear more needy (and eligible for more programs).

5. *Phone charging.* In the expenditure section of the survey, the survey asked how many times in the last two weeks the respondent charged his phone for money. This corresponds to taking one's phone to a kiosk with electricity where one pays a small fee to recharge the battery, a common and routine expense for many Liberians, without stigma and possibly not very memorable. 38% of our sample had a mobile phone at the endline, and 38% reported charging a phone in the last two weeks.

6. *Video Club Attendance.* In the expenditure section of the survey, the survey asked how many times in the last two weeks the respondent went to a video club. These clubs are private businesses where one can go to watch a movie, television show, or football match for a small fee. This is a popular and socially acceptable pastime, as most Liberians do not have electricity or home entertainment. Salience was unclear but likely greater than phone charging.

These behaviors also exhibited diversity in program emphasis. Some, like stealing and marijuana use, were highly emphasized in the STYL therapy, while others like video club and phone charging were not.

F.2 Validator staff

Eight local staff performed validations over the two years of data collection. We selected validators from the study's qualitative research staff. These people typically began as survey enumerators, but displayed such skill and rapport with the subjects that we hired and trained them to conduct a separate qualitative research component: longitudinal, formal, open-ended interviews with a different subsample of subjects. All conducted the qualitative validation when they were not working on the formal open-ended interviews.[20]

Each validator received at least 10 days of training on the methods, including both classroom learning and extensive field training. We trained more qualitative researchers than were needed for the exercise. Those who exhibited superior performance during the trainings were selected as

---

[20]All but one were men, and all had a high school education. Two of the men completed roughly half the validations with the remainder doing roughly 10 to 20% each. To find these validators, we trained roughly two to three times the number of people needed from the pool of research staff, selecting only those with the most natural questioning and rapport-building skills for the validation exercise.

validators. The aim of the training was to develop and refine trainees' skills in acquiring informed consent, building rapport with respondents, collecting and recording data, and analytical reasoning. Trainings were held for eight hours each day and, over the course of 10 days, transitioned gradually from exclusive classroom learning to field trainings with short debriefing sessions. Field trainings provided trainees with opportunities to practice the skills and techniques they had learned.

Like any qualitative study, we believe staff recruitment and training to have been among the most important tasks and also the largest start-up cost of this method.

F.3 Approach

For each respondent, validators tried to determine whether the respondent had engaged in any of the measured behaviors, even once, in the two weeks preceding the respondent's survey date, as the survey asked about behaviors occurring during the two weeks prior to the survey. We found it optimal for validators to visit each respondent four times, on four separate days, with each visit or "hangout session" lasting approximately three hours. The validator aimed to begin hanging out the day after subjects completed their quantitative surveys and to conduct all four visits in the days following the respondent's endline survey date.

Validators deliberately avoided the feeling of a formal interview and would typically accompany respondents as they went about their business.[21] Validators sometimes took notes during visits, but only in isolated areas out of sight from the respondent.[22] The idea follows from basic principles of ethnography, which seeks to study subjects in their natural settings, similar to those the researcher hopes to generalize about. The intent is to reduce the sense of being in an experimental situation, which ethnographers perceive as creating bias.

The main approach was to engage in casual conversation on a wide range of topics, including the six target topics/measures. The target topics were raised mainly through indirect questions while informally chatting. For example, validators typically started conversations with discussions of family. This was both customary among peers in Liberia and a sign of respect and interest in respondents' lives. It was also a stepping stone for discussing the target behaviors—either because the validator can discuss an issue in their family (someone engaging in one of the activities) or how the respondent's family feels about their current lifestyle and circumstances.

In general, validators found it helpful to tell respondents stories or scenarios about another person or themselves, related to the target measures, then steer the conversation to get information about how respondents had behaved in similar situations, eventually discussing the past two weeks. Validators were careful to present these behaviors and incidents in a non-stigmatized light, for instance by discussing a friend who stole in order to get enough to eat, or how they themselves had periods of homelessness or used drugs and alcohol. Validators found these personal stories (all of which were truthful) and genuineness were essential to building rapport and trust.

Validators might hold these conversations once or twice over the three hours, spending perhaps twenty or thirty minutes in conversation each time, to avoid unnaturally long or awkward conversations. The validator spent the remainder of the three hours in the general vicinity, observing

---

[21]On the first visit validators would obtain verbal consent. We designed the consent script to be informal, and explained that the goal of hanging out with the respondent was to talk about some of the same things they discussed in the survey. In addition to this verbal consent, the formal consent form that preceded the recent survey said that qualitative staff may come and visit them again to gather more information.

[22]e.g. in a toilet stall or teashop. If validators were unable to find a secluded area in which to take notes, they sometimes recorded information in their cell phones, pretending to send a text message.

respondents engaging in their daily activities. This could involve taking a rest in the shade or in a tea shop (as is common) or engaging others in conversation. Validators would also try to talk casually with the respondent's friends, relatives, or neighbors to learn about him (although we considered information from these second-hand sources as insufficient to support a conclusion about the respondents' behaviors, but merely as supporting information).

We found that building a rapport with participants in a short space of time was crucial. To develop trusting and open relationships, validators used techniques, including becoming close to respected local community and street leaders, eating meals together, sharing personal information about themselves, assisting subjects with daily activities, and mirroring participants' appearances and vernacular, as appropriate. In addition, validators tried to maintain neutrality and openness while discussing potentially sensitive topics. For instance, conveying—through stories or otherwise—that illicit behaviors were not perceived negatively, allowed respondents to feel comfortable sharing their involvement in such activities. Validators did not lie to or deceive respondents, however.

Overall, this approach—trust-building, spending time together over the course of several days, assuming the role of an "insider," attempting to obtain admission or discussion of the behavior, clandestine but fairly immediate note-taking, and (as discussed below) close examination of the evidence for each respondent with the investigators—was designed to counter the observer bias and selective recall that concern participant observation.[23] Developing a rapport with respondents, spending time to develop a relationship, and obtaining insider status are considered central to obtaining more honest and valid responses (Baruch, 1981; Bryman, 2003; Fox, 2004). We are not aware of any study, however, that has quantitatively tested this proposition.

F.4 Validation sampling and non-response

In each endline survey round we randomly selected study respondents to be validated, stratified by treatment group.[24] Table F.1 describes the samples selected for validation in each survey round over the course of the study. In total, we randomly selected 7.4% of all surveys, 297 in total, for validation.

We found 240 (81%) of the 297.[25] This attrition is an identification concern, but there is little evidence of biased attrition. Excess validation attrition (those who were surveyed but not validated) was not robustly associated with baseline characteristics (see Appendix A.3).

Statistical power. In order to minimize the confidence intervals surrounding any treatment-measurement error correlation, we chose the sample size that maximized the number of interviews

---

[23]For general discussions of validity in qualitative methods, see Wilson (1977); LeCompte and Goetz (1982); Power (1989).

[24]For each pair of survey rounds, study participants were randomly divided into blocks (e.g. 1, 2, 3, 4), and block 1 study participants were surveyed before block 2, and block 2 before block 3, etc. Within each block we randomly selected validation subjects using a computer-generated uniform random variable. The selection was performed without replacement in a given pair of survey rounds (e.g. the short-term endline surveys in a given phase), but sampling was performed with replacement across survey rounds. Twenty subjects were validated in more than one round.

[25]We could not find 15 for even the endline survey. We could not validate a further 42 because they were difficult to find even immediately after the survey or (more commonly) because they lived a long distance away. In general, we surveyed respondents who had moved far out of Monrovia, but we were unlikely to validate them because of the time and expense and opportunity cost.

Table F.1: Validation sample, totals and attrition

Phase	Round	Surveys Target #	Validation Selected	Validation Validated	Validation Unfound at endline	Reason for no validation data Unfound for validation	Reason for no validation data % validated (all)	Reason for no validation data % validated (treatment)	Reason for no validation data % validated (control)
	3-week	100	0						
	5-month	100	24	18	2	4	75%	75%	75%
1	7-month	100	24	12	1	11	50%	50%	50%
	12-month	100	10	6	3	1	60%	63%	50%
	13-month	100	10	8	2	0	80%	86%	67%
	3-week	398	26	24	0	2	92%	94%	89%
2	5-week	398	27	17	0	10	63%	68%	40%
	12-month	398	28	25	2	1	89%	86%	100%
	13-month	398	44	38	1	5	86%	85%	91%
	3-week	501	0						
3	5-week	501	0						
	12-month	501	35	31	2	2	89%	89%	88%
	13-month	501	69	61	5	3	88%	88%	88%
All		4096	297	240	18	39	81%	81%	80%

Notes: The proportion selected in each round was principally a function of logistical feasibility (e.g. number of available staff), and in some cases none were selected. As procedures became more familiar and staff more experienced, more could be done over time. The percentage validated in the treatment group includes any treatment (cash, CBT, or both).

we felt qualified validators could manage logistically.[26] Post hoc calculations of statistical power confirm the estimates we made at the design stage. With a sample of 240, we can detect general over- or under-reporting greater than 17% of the survey mean (14% of the "true" validated mean).[27] Because each treatment arm is a subsample, however, we cannot precisely measure the effect of treatment on misreporting—it is difficult to detect effects greater than 33% of the survey mean (28% of the validated mean). Thus we are principally interested in the sign and magnitude of the treatment effect on misreporting by treatment group.

F.5 Coding validated data

Validators were unaware of the respondents' survey responses, and formed their own opinions (based on the evidence collected) about whether respondents engaged in the six activities during the time period captured by the quantitative survey. Every coding recommendation was then discussed with and vetted by one of the authors.

A core part of the validator training included logical reasoning, supporting reasoning with evidence, and writing this down in a clear and structured manner. After each visit, validators made written notes about the relevant data collected, including evidence to support their conclusions, on a standardized form. At the conclusion of the four visits, the validator coded six indicators, one for each behavior, where "1" meant that he had relatively direct evidence that the respondent engaged in the behavior during the recall period, and "0" otherwise.[28]

Validators recorded an average of 1.35 "major" pieces of evidence per respondent per behavior to support their coding decision sheets. This was typically the most persuasive piece or pieces of

---

[26]In general, the validation sample was a balanced subsample of the full sample. Power calculations, based on roughly the first 60 validator interviews, indicated that there was a modest degree of underreporting of all behaviors, sensitive and non-sensitive, but that the correlation between treatment status and measurement error was uncertain—across outcomes it varied in sign and magnitude, but was about zero on average. Thus the chief advantage of maximizing the sample conditional on time available was to shrink the confidence interval to build confidence in our method and the main outcomes of interest. Further validation was mainly limited by the number of validators we felt could be trained and supervised.

[27]We calculated this minimum detectible effect (MDE) using a two-sided hypothesis test with 80% power at a 0.05 significance level, using baseline and block controls when calculating the R-squared statistic. We calculated an MDE for both the 0–2 expenditures index and the 0–4 sensitive behaviors index. The expenditures index had a mean of .82 in the survey and an MDE of .13 for general over- and under-reporting and .29 for a treatment effect on misreporting. The sensitive behaviors index had a mean of 1.12 in the survey and an MDE of .2 for general over- and under-reporting and .36 for any treatment effect on misreporting. We estimate that doubling the sample size would have increased power by about a third.

[28]Over the course of the exercise, different measures offered different experiences and lessons. Because of its relative frequency and visibility, we suspect marijuana use was the easiest to directly observe. But validators found other behaviors straightforward to discuss in conversation. In the survey and (especially) the validation, phone battery charging led to the most confusion—in particular, did simply charging one's phone count, or did only paying to charge one's phone count? Paid charging was the focus of the survey question (it appeared in an expenditure survey module), but we were concerned that the validators would use a more expansive definition. We attempted to mitigate such differences through trainings and regular discussions on the coding.

Homelessness also proved somewhat challenging to measure and validate, as we discovered its definition is subjective. Circumstances arose that were somewhat ambiguous, such as having no home of one's own but regularly sleeping on a friend's floor or in an acquaintance's market stall. To account for the potential variability in perceptions of homelessness, validators were instructed to include as much information as possible about respondents' living situations in their summary reports. The authors then worked with validators to code a somewhat broad definition of homelessness that included any ambiguous circumstances. Prior to analysis, it was not clear whether survey respondents applied the same definition, and hence we err on the side of finding underreporting in the survey.

evidence rather than all evidence collected.[29] Table F.2 reports evidentiary methods by behavior. In general, the validators used some form of direct or indirect questioning—a direct admission of the behavior or persuasive statements that they did not engage in the behavior. The validators only witnessed or found direct evidence of the behavior in a fifth of cases, or had third party verification in about 6% of cases. In any event, witnessing or third party verification were not sufficient evidence for a final coding. For instance, witnessing had to be followed by questions confirming that the respondent also engaged in the behavior in the two weeks prior to the survey. This accounts for most of the cases where there was more than one piece of evidence highlighted.

In general, the patterns of evidence are fairly commonsensical. Witnessing is limited to observable behaviors such as marijuana, gambling, homelessness, and phone charging. Stories and scenarios where the respondent is invited to comment or discuss are especially common for the most sensitive subject, stealing. Indirect questioning is most common for everyday topics such as homelessness ("Is this your house?") and phone charging ("I need to charge my phone. Where do you usually charge yours?").

F.6   Limitations of the approach

While we think, based on our experiences, that this validation exercise gave enough time to gather detailed, accurate information and fostered trust and frankness, there are nonetheless limitations to this approach.

1. POTENTIAL DISRUPTION. The presence, and interactions and conversations with the validators may be intrusive and might disrupt respondents' daily activities, thereby altering the findings. To mitigate this risk, validators wore clothes that would blend in with their respondent's environment, and typically accompanied and assisted respondents in their activities as appropriate (e.g. helping a scrap metal collector scavenge).

2. DIFFERENCES IN RECALL PERIODS. The validation occurred after the time period about which the survey questions had asked, and validators or respondents could have made errors about the relevant window of time (e.g. homelessness could have been observed the week after the survey, and inferred to the time of the survey incorrectly). This is most likely a source of random measurement error.

3. INCONSISTENT QUESTIONS. The survey and validation questions might have been interpreted differently, making it difficult to compare results. As discussed above, phone charging and homelessness proved somewhat difficult to measure consistently. We used close consultations and reviews of the data, and focus groups with survey and validation staff, to maximize consistency.

4. REVERSE HAWTHORNE EFFECT. Training validators to look for certain behaviors could lead them to overreport those behaviors (akin to the problem of "when you have a hammer everything looks like a nail"). This reverse Hawthorne effect would probably be more of a risk if the validation method relied on passive observation. Rather, validation involved active discussion and (usually) a direct admission of the behavior. Also, one of the authors reviewed and discussed the evidence for every subject with the validator.

---

[29]We do not have complete paper records of all evidence collected, and so the 1.35 pieces of evidence is probably an understatement of the full amount of evidence.

Table F.2: Evidentiary methods reported by validators, by behavior

Main evidence techniques	Potentially sensitive behaviors				Expenditures	
	Steal	Marijuana	Gamble	Homeless	Video	Phone
	(1)	(2)	(3)	(4)	(5)	(6)
Avg. pieces of evidence	1.1	1.3	1.1	1.7	1.0	1.2
Obs. (All)	240	240	239	240	239	240
Direct question	36%	35%	38%	5%	32%	1%
Indirect question	28%	46%	42%	62%	59%	92%
Story / scenario	36%	6%	13%	12%	2%	1%
Witnessed / found evidence	3%	31%	9%	62%	5%	18%
Third party account	3%	6%	4%	21%	0%	0%
Other / unclear	3%	9%	6%	13%	6%	5%
Obs. (Coded "did not engage" in behavior)	191	118	170	190	93	125
Direct question	38%	44%	39%	5%	34%	0%
Indirect question	26%	46%	44%	60%	58%	98%
Story / scenario	37%	7%	15%	12%	3%	2%
Witnessed / found evidence	2%	3%	1%	65%	2%	1%
Third party account	3%	10%	4%	24%	0%	1%
Other / unclear	2%	1%	1%	14%	4%	0%
Obs. (Coded "did engage" in behavior)	49	122	69	50	146	115
Direct question	29%	25%	36%	4%	30%	2%
Indirect question	33%	46%	38%	70%	60%	86%
Story / scenario	33%	5%	9%	10%	1%	0%
Witnessed / found evidence	10%	59%	28%	52%	7%	37%
Third party account	4%	2%	4%	8%	0%	0%
Other / unclear	8%	17%	17%	6%	8%	10%

Notes: Direct questions imply the validator asked the respondent directly about his engagement in the activity. Indirect questions imply the validator brought up the subject in general conversation (Where do you live? What do you do to make money?). Stories and scenarios are a form of indirect questioning where the respondent is invited to comment. Witnessing or found evidence implies the validator saw the respondent engaging in the activity in question or found physical evidence that the respondent recently engaged in the activity. Third party accounts imply the validator asked the family and friends of the respondent whether or not he engaged in the activity. Other or unclear methods include a handful of cases of unprompted information from the respondent, and also cases where the behavior could be inferred from other knowledge. Mainly it implies that coding was inconclusive or incomplete but is likely a form of questioning.

5. INCREASING SOCIAL DESIRABILITY BIAS. In principle the participant observation method, by building rapport, could lead to a different source of measurement error by (for example) increasing social desirability bias. Our strong sense is that the opposite is true, that trust and rapport reduced the bias, but this is a subjective interpretation and not independently verifiable.

6. CONSISTENCY BIAS. In principle, respondents could recall their survey response and try to remain consistent despite trust-building. This could motivate randomizing the order of validation and survey in the future.

7. NON-BLINDED VALIDATORS. The researcher is not immune from bias in qualitative research (LeCompte and Goetz, 1982; LeCompte, 1987). We are especially concerned with any bias correlated with treatment. While validators weren't given the subject's treatment status, it's possible and even likely that this could come up during the extended conversations. Thus there is a danger that the validators' biases will be correlated with treatment. The trust-building and preference for direct admission of the behavior was intended to mitigate this risk, but it still remains.

Most importantly, it seems unlikely that validators would commit most of these errors differentially across study arms. Misreporting correlated with treatment is still a risk under the consistency bias and non-blinded limitations, but the in-depth focus on a handful of questions, time invested, and trust-building is designed to counteract these biases as much as possible. If so, the qualitative validation method may be most useful at building confidence estimated treatment effects.

Finally, like any qualitative work, this is not an off-the-shelf tool. To select and refine the variables, recruit and train validators, and monitor quality of the data requires the researcher to have some familiarity with the context and population and at least basic experience in qualitative data collection.

F.7   Replicability of the approach

There are three reasons to think that this method could be replicated in other developing country field experiments and observational analysis using surveys. First, the expertise needed to implement the method effectively exists in most countries. Indeed, it should be considerably simpler to implement outside than inside Liberia. After fourteen years of civil war, and with one of the lowest human development indices in the world, Liberia has very low local research capacity, even compared to other poor and post-conflict states.

Second, most social scientists are nearly as well prepared to design and implement the approach as they are a new survey instrument or measure. Like any measure or method, it takes local knowledge, care, and extensive pretesting to develop a credible approach, and can benefit from someone with expertise in the subject area. In our case, one of the field research managers had some background in qualitative work and quality assurance, which we believe improved the quality of training and selection of the validator staff.

Third, the cost of the data collection is not necessarily large relative to many field experiments or large-scale panel surveys. In this instance, the fixed cost of startup was primarily in the recruitment and training of the small number of validators—approximately 2 to 3 weeks of work. We estimate

the marginal cost of validation was roughly \$80 per respondent, mainly in wages and transport. By comparison, the marginal cost of surveying a respondent was roughly \$70.[30]

While this method is considerably more expensive than survey experiments, it is more in line with the depth and cost of commonplace efforts to improve consumption measurement through the use of diaries physical measurement.[31] For crucial measures in large program evaluations, or for statistics informing major policies, the cost is small relative to the intervention, larger study, or larger purpose. For instance, as a proportion of total expenditures on the study, this validation exercise cost under 3% of all research-related costs, and less than 1–2% of program plus research costs.

F.8 Further results: Misreporting levels

Table F.3 reports our proxy of survey over-reporting: the simple survey-validation differences, with p-values from a t-test of the difference from zero. Negative values indicate survey under-reporting, assuming the validator measure is more accurate of course. As noted above, we have the statistical power to detect differences greater than about 17% of the survey mean.

Overall, gambling seems to be slightly underreported in every treatment arm, and highly under-reported by men in the control and cash only groups. For instance, 33% of the cash only group admitted to gambling during validation, compared to 13% during the survey. Some of this underreporting could be due to ambiguous behaviors being coded as gambling in validation interviews but not in the survey. But the fact that underreporting is smaller in the therapy arms suggests that the underreporting is not an artifact of different definitions, but rather reflects a strategic response to treatment status.

If we look at stealing, marijuana use, and homelessness, however, none of the survey-validation differences are statistically significant. There is possibly some slight underreporting of drug use and slight over-reporting of stealing, but the magnitudes are generally small in the sense that they are less than 10% of the survey means reported in Table 9. The sample size is small, however, and so many of these differences are not precisely estimated.

We see much stronger evidence of underreporting of expenditures in the survey. The difference for both expenditures is -0.27 in the full sample (Table F.3, Column 6). This difference is large—about a third of the survey mean reported in Table 9. Expenditure underreporting is largest for the video club measure, but both expenditures appear to be underreported. Interestingly, the mean differences appear to be smaller and less statistically significant if the men received one of the treatments. We return to these differences across treatment arms below.

F.9 Further results: Adjusted treatment effects

We estimate the effect of each treatment on survey over-reporting, in Table F.4. These estimates effectively take the simple survey-validation differences in Panel A of Table 10 and estimate the difference across treatment arms, adjusting for baseline covariates as well as block fixed effects. We use these to calculate an adjusted treatment effect.

---

[30]Both figures were driven by the fact that it typically took one to two days of searching to find each respondent for surveying, plus the time to survey itself. Both surveying and validating in Liberia were expensive by the standards of household surveys, largely because of the cost of operating in a fragile, post-conflict state and the great difficulties in tracking such an unstable population.

[31]In one extreme example, in the India NSS consumption survey, enumerators physically measure the volume of all food consumption Group (2003).

Table F.3: Survey over-reporting, estimated by the mean difference between survey and validation measures (y^{Δ})

	Potentially sensitive behaviors					Expenditures		
	All (0-4)	Steal	Marijuana	Gamble	Homeless	All (0-2)	Video	Phone
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample	-0.10	0.02	-0.03	-0.11	0.02	-0.27	-0.19	-0.08
	0.17	*0.57*	*0.24*	***0.00***	*0.45*	***0.00***	***0.00***	***0.00***
Control group	-0.07	0.03	-0.02	-0.12	0.03	-0.50	-0.29	-0.22
	0.64	*0.57*	*0.71*	*0.09*	*0.60*	***0.00***	***0.00***	***0.00***
Therapy only	-0.04	0.02	0.00	-0.07	0.02	-0.17	-0.13	-0.04
	0.80	*0.77*	*1.00*	*0.29*	*0.77*	*0.08*	*0.07*	*0.53*
Cash only	-0.29	-0.02	-0.05	-0.20	-0.03	-0.23	-0.18	-0.05
	0.04	*0.80*	*0.37*	***0.00***	*0.42*	***0.03***	***0.03***	*0.32*
Therapy + cash	0.02	0.03	-0.05	-0.03	0.06	-0.19	-0.16	-0.03
	0.91	*0.57*	*0.37*	*0.66*	*0.25*	***0.01***	***0.02***	*0.48*
Observations	239	238	238	238	239	239	238	239

Notes: Columns 1 to 8 report the simple mean differences in the survey and validation measures for the full sample and for each treatment arm, along with p-values for as t-test of whether the mean is different from zero. We bold p-values ≤ 0.05.

First, the results imply that the adjusted treatment effect of therapy and cash on sensitive behaviors overall is no lower than what we estimate with self-reported survey data, and may even be larger (Column 1). This holds true for each of the individual sensitive behaviors, save marijuana use. Despite the large standard errors introduced by the small validation sample, the adjusted treatment effect on all sensitive behaviors is larger and significant at the 1% level.

Meanwhile, the underreporting of gambling does not have a statistically significant association with treatment. Those who received cash alone underreported gambling to the surveyors more often than control group members, and so the measurement error in gambling is probably a combination of a general desirability bias as well as one correlated with treatments. A larger sample size would be needed to separate these more precisely.

In contrast, the slight underreporting of expenditures behaviors in the survey (seen in Table F.3 above) implies that the short term increase in survey-based expenditures due to cash could be due to measurement error correlated with treatment. The adjusted treatment effect of therapy plus cash is generally negative but not statistically significant (Column 6). We see a similar pattern with another expenditure-related item, homelessness, in Table F.4—the survey-reported decline in homelessness tends to disappear with adjustment.

References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Table F.4: Estimates of adjusted treatment effects by outcome and treatment

Adjusted ATE	Potentially sensitive behaviors					Expenditures		
	All (0-4)	Steal	Marijuana	Gamble	Homeless	All (0-2)	Video	Phone
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cash only	0.121	-0.014	0.058	0.072	0.001	-0.205	-0.079	-0.132
	[.196]	[.088]	[.074]	[.093]	[.077]	[.143]	[.117]	[.079]*
Therapy only	-0.190	-0.041	-0.031	-0.131	0.007	-0.334	-0.167	-0.172
	[.202]	[.082]	[.066]	[.104]	[.087]	[.148]**	[.121]	[.090]*
Therapy and cash	-0.516	-0.122	-0.048	-0.198	-0.156	-0.239	-0.125	-0.120
	[.194]***	[.075]	[.071]	[.095]**	[.082]*	[.137]*	[.113]	[.077]
Observations, survey/validation	3765 / 239	3764 / 238	3762 / 238	3763 / 238	3765 / 239	3763 / 239	3761 / 238	3759 / 239

Notes: The survey-based ATE estimates pool all survey rounds and regress each outcome on treatment indicators and block fixed effects. Standard errors are robust and clustered by individual. Estimates of the bias from treatment come from a regression of the difference in the survey and validation measures on an indicator for treatment arms, controlling for block fixed effects and each endline round. Standard errors are robust and clustered by block. The difference is an estimate of the true treatment effect after adjusting for observed bias. It is calculated as the linear difference of the estimates and the standard error is calculated via bootstrapping.

*** p<0.01, ** p<0.05, * p<0.1

Aron, A. (2007). The neural basis of inhibition in cognitive control. *Neuroscientist 13*(3), 214–28.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of adhd. *Psychological Bulletin 121*(1), 65–94.

Baruch, G. (1981). Moral tales: parents' stories of encounters with the health professions. *Sociology of Health & Illness 3*(3), 275–295.

Blattman, C. and J. Annan (2015). Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State. *forthcoming in American Political Science Review*.

Blattman, C., N. Fiala, and S. Martinez (2014). Generating skilled employment in developing countries: Experimental evidence from Uganda. *Quarterly Journal of Economics 129*(2), 697–752.

Bryman, A. (2003). *Quantity and quality in social research.* New York: Routledge.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Draca, M. and S. Machin (2015). Crime and Economic Incentives. *Annual Review of Economics 7*, 389–408.

Eigsti, I.-M., V. Zayas, W. Mischel, Y. Shoda, O. Ayduk, M. Dadlani, M. Davidson, A. J. Lawrence, and B. Casey (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science 17*(6), 478–84.

Ersche, K., A. Turston, S. Pradhan, E. Bullmore, and T. Robbins (2010). Drug addiction endophenotypes: impulsive versus sensation-seeking personality traits. *Biological Psychiatry 68*(8), 770–3.

Fafchamps, M., D. J. McKenzie, S. Quinn, and C. Woodruff (2014). When is capital enough to get female microenterprises growing? Evidence from a randomized experiment in Ghana. *Journal of Development Economics 106*(1), 211–226.

Fox, R. C. (2004). Observations and reflections of a perpetual fieldworker. *The ANNALS of the American Academy of Political and Social Science 595*(1), 309–326.

Group, N. E. (2003). Suitability of different reference periods for measuring household consumption. results in pilot survey. *Economic and Political Weekly 38*(4), 25–31.

Holt, C. and S. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Jamison, J. and D. Karlan (2011). Measuring preferences and predicting outcomes. *Working paper*.

Karlan, D., R. Knight, and C. Udry (2015). Consulting and capital experiments with microenterprise tailors in ghana. *Journal of Economic Behavior and Organization 118*, 281–302.

LeCompte, M. D. (1987). Bias in the biography: Bias and subjectivity in ethnographic research. *Anthropology & Education Quarterly 18*(1), 43–52.

LeCompte, M. D. and J. P. Goetz (1982). Problems of reliability and validity in ethnographic research. *Review of educational research 52*(1), 31–60.

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *American Economic Review 80*(2), 319–323.

Nelson, C. A. e. a. (2007). Cognitive recovery in socially deprived young children: The bucharest early interventino project. *Science 318*(5858), 1937–1940.

Power, R. (1989). Participant observation and its place in the study of illicit drug abuse. *British Journal of Addiction 84*(1), 43–52.

Udry, C. (2010). The Economics of Agriculture in Africa: Notes on a Research Program. *African Journal of Agricultural and Resource Economics 5*(1), 284–299.

Wilson, S. (1977). The use of ethnographic techniques in educational research. *Review of educational research 47*(1), 245–265.